Universidade de Aveiro Departamento de Biologia
2018

Olena Oleksandrivna Kutsenko

Regulation of *MCL1* 3’ UTR-APA-derived mRNA isoforms

Regulação das isoformas de mRNA de *MCL1* derivadas de 3’ UTR-APA

DECLARAÇÃO

Declaro que este relatório é integralmente da minha autoria, estando devidamente referenciadas as fontes e obras consultadas, bem como identificadas de modo claro as citações dessas obras. Não contém, por isso, qualquer tipo de plágio quer de textos publicados, qualquer que seja o meio dessa publicação, incluindo meios eletrónicos, quer de trabalhos académicos.

Universidade de Aveiro Departamento de Biologia
2018

Olena Oleksandrivna Kutsenko

Regulation of *MCL1* 3' UTR-APA-derived mRNA isoforms

Regulação das isoformas de mRNA de *MCL1* derivadas de 3' UTR-APA

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular, realizada sob a orientação científica da Doutora Isabel Pereira de Castro, Investigadora Pós-Doutoramento no Instituto de Biologia Molecular e Celular da Universidade do Porto (IBMC)/ Instituto de Investigação e Inovação em Saúde (i3S), da Professora Doutora Maria Alexandra Marques Moreira Mourão do Carmo, Investigadora Principal no Instituto de Ciências Biológicas de Abel Salazar (ICBAS) da Universidade do Porto, e líder do grupo "Gene Regulation" no Instituto de Biologia Molecular e Celular (IBMC)/ Instituto de Investigação e Inovação em Saúde (i3S) da Universidade do Porto, e da Doutora Maria de Lourdes Gomes Pereira, Professora Associada com Agregação ao Departamento de Ciências Médicas da Universidade de Aveiro.

O trabalho desenvolvido nesta dissertação foi financiado por fundos do Norte-01-0145-FEDER-000008 – "Porto Neurosciences and Neurologic Disease Research Initiative" no I3S, apoiado pelo NORTE 2020 (Programa Operacional Regional do Norte 2014/2020), sob o Acordo de Parceria - PORTUGAL 2020, através do Fundo Europeu de Desenvolvimento Regional (FEDER).

O júri

Presidente	Doutor Mário Guilherme Garcês Pacheco Professor Associado com agregação ao Departamento de Biologia da Universidade de Aveiro
Arguente Principal	Doutora Raquel Monteiro Marques da Silva Professora Auxiliar Convidada do Departamento De Ciências Médicas da Universidade de Aveiro
Orientadora	Doutora Isabel Pereira de Castro Investigadora Pós-Doutoramento no Instituto de Biologia Molecular e Celular (IBMC)/ Instituto de Investigação e Inovação em Saúde (i3S) da Universidade do Porto

Durante a realização deste trabalho foram realizadas as seguintes comunicações orais e em poster:

Comunicações em formato de Poster:

Kutsenko, O.; Pereira-Castro, I.; Moreira, A.; An Innovative Method For Polyadenylation Signal Targeting By CRISPR/Cas9 In Difficult To Transfect Cells. XI ENEBIOQ, 23-26 th March 2018, Associação dos Industriais da Construção Civil e Obras Públicas (AICCOPN), Porto, Portugal.

Comunicações orais:

Pereira-Castro I, **Kutsenko O**, Curinha A, Neilson JR, Moreira A.; *MCL1* is regulated by post-transcriptional mechanisms in human T cells. Annual Meeting of the Portuguese Society of Genetics, 14-15 th June 2018, Instituto de Inovação e Investigação em Saúde (i3S), Porto, Portugal.

Agradecimentos

Estou muito grata à Alexandra Moreira por me ter acolhido no grupo. Agradeço muito a minha orientadora, Isabel Castro, que sempre se disponibilizou para me ajudar e aconselhar ao longo do meu percurso, dentro dos possíveis. Ainda, agradeço à Joana Wilton e à Marta Oliveira pelo apoio que me têm dado.

palavras-chave Poliadenilação alternativa (APA); Região 3’ não traduzida (3' UTR); Myeloid Cell Leukemia 1 (MCL1); ensaios de localização; CRISPR / Cas9; transdução lentiviral; knock-out (KO); INTS9; CPSF1.

resumo

A poliadenilação alternativa (APA) é uma etapa de processamento de pre-mRNA em que isoformas alternativas de mRNA são formadas com base na escolha de um sinal de poliadenilação (pA) sobre o outro. Quando confinada à Região 3' não-Traduzida (3' UTR), a APA gera mRNAs que variam somente no comprimento das suas 3' UTRs (3'UTR-APA). Vários estudos têm demonstrado o papel da APA na regulação da expressão genética. Nos mamíferos, cerca de 70% de genes são sujeitos a APA, aumentando substancialmente a diversidade do transcriptoma. APA é relevante tanto em processos fisiológicos como patológicos, modulando a estabilidade, localização e eficiência da tradução dos transcritos. Trabalhos anteriores do grupo demonstraram que o gene Myeloid Cell Leukemia 1 (*MCL1*) tem dois sinais de poliadenilação ativos em células T humanas, denominados pA1 e pA2, e que a isoforma pA2 é susceptível à regulação pós-transcripcional pelo miR-17 no citoplasma. O produto proteico comum às duas isoformas de APA – Mcl-1, é um membro anti-apoptótico da família Bcl-2, crucial tanto no câncro como na sobrevivência das células T. Um dos objectivos desta dissertação consistiu em abordar a influência das duas isoformas de APA de *MCL1* na localização subcelular de Mcl-1 e em verificar se o miR-17 tem um impacto no seu nível de produção proteica. Deste modo, os seguintes construtos de expressão contendo a sequência codificante de *MCL1* (CDS), seguida pela sequência da 3' UTR curta ou longa, foram transfetados em células HeLa; pEGFP *MCL1* CDS pA1, pEGFP *MCL1* CDS pA2 e *MCL1* CDS pA2∆miR17, e a localização de Mcl-1 foi verificada através da microscopia confocal. Pretendíamos ainda otimizar o método da deleção dos sinais pA em células difíceis de transfetar (Jurkat E6.1), através do uso de construtos CRISPR/Cas9 do tipo *all-in-one* e transdução lentiviral. Para além de planear a criação de linhas com um dos sinais pA de *MCL1* deletado: *MCL1*∆pA1 e *MCL1*∆pA2, respectivamente, o nosso objectivo consistiu em criar linhas estáveis *knock-out* (KO) para ambos alelos em dois genes envolvidos no processamento de RNA: INTS9 e CPSF1, para estudar a sua função na APA e na expressão de *MCL1*. Os ensaios de localização mostraram que Mcl-1 derivado da isoforma pA1 é predominantemente expresso na mitocôndria, enquanto Mcl-1 derivado da isoforma pA2 tem uma localização ubíqua. Além disso, a quantificação da intensidade de EGFP mostrou que a isoforma pA1 produz ~ 3 vezes mais proteína que a pA2 e que o miR-17 não é responsável pela baixa expressão de Mcl-1. Por outro lado, a genotipagem das populações de E6.1 submetidas a CRISPR/Cas9 revelou deleções em alelos de quase todas as condições. De seguida, a heterozigotia foi detectada em clones isolados por *single-cell sorting*. Em conclusão, os estudos de localização provam que a regulação da expressão genética mediada pela 3' UTR modula tanto o destino como a distribuição subcelular proteica. As abordagens CRISPR/Cas9 foram bem-sucedidas, contudo, a homozigotia ainda está por alcançar, para permitir a realização de ensaios funcionais. Em suma, estes resultados ajudam a esclarecer a regulação de isoformas de mRNA derivadas de 3' UTR-APA de MCL1, proporcionando-nos oportunidades para aprofundar o conhecimento sobre os mecanismos e os fatores moleculares subjacentes.

keywords Alternative polyadenylation (APA); 3’ Untranslated Region (3’ UTR); Myeloid Cell Leukemia 1 (*MCL1*); localization assays; CRISPR/ Cas9; lentiviral transduction; *knock-out* (KO); INTS9; CPSF1.

abstract

Alternative polyadenylation (APA) is a pre-mRNA processing step in which mRNA isoforms with alternative 3' ends are formed based on the choice of one polyadenylation (pA) signal over the other. When confined to the 3' Untranslated Region (3' UTR), APA generates mRNAs varying only in the length of their 3' UTRs (3'UTR-APA). Various studies have demonstrated a role for APA in regulating about 70% of mammalian genes, increasing immensely the transcriptome diversity. APA is relevant in both physiological and pathological processes, primarily by modulating transcripts stability, localization and translational efficiency. From our group previous work, it was shown that the Myeloid Cell Leukemia 1 (*MCL1*) gene has two active pA signals in human T cells that were named pA1 and pA2 and that the pA2 isoform is susceptible to post-transcriptional regulation by miR-17 in the cytoplasm. The APA isoforms common protein product – Mcl-1, is an anti-apoptotic member of the Bcl-2 protein family, crucial in both cancer and T cell survival. One of the aims of this thesis consisted in assessing the influence of the two *MCL1* APA isoforms on subcellular distribution of Mcl-1 and verify if miR-17 has an impact on Mcl-1 protein production. In order to do so, the following expression constructs featuring *MCL1* coding sequence (CDS) followed by either the short or the long 3' UTR sequence were transfected in HeLa cells; pEGFP *MCL1* CDS pA1, pEGFP *MCL1* CDS pA2 and MCL1 CDS pA2ΔmiR17, and Mcl-1 localization was verified by live-cell confocal microscopy. In addition, we aimed to optimize a method for pA signal deletion in difficult to transfect cells (Jurkat E6.1), by using all-in-one CRISPR/Cas9 constructs and lentiviral transduction. Aside from planning to generate *MCL1* lacking either the proximal (*MCL1*ΔpA1) or the distal (*MCL1*ΔpA2) signal, it was our objective to generate stable E6.1 double-allele knock-outs (KO) for two RNA processing-involved genes: INTS9 and CPSF1 and assess their effect on *MCL1* APA. The localization assays showed that the pA1 isoform-derived Mcl-1 was predominantly expressed in the mitochondria while the pA2/ pA2ΔmiR17 isoforms-derived Mcl-1 had a ubiquitous localization. Furthermore, the quantification of EGFP intensity showed that the pA1 isoform produced ~3-fold more protein than pA2, and that miR-17 was not responsible for pA2 isoform-derived Mcl-1 low expression. On the other hand, genotyping of CRISPR/Cas9-subjected bulk E6.1 populations has rendered deletion-positive alleles in almost all conditions. Then, heterozygosity was detected in single-cell sorting isolated clones. In conclusion, the localization studies prove that the 3' UTR-mediated regulation of gene expression reaches into protein fate modulation, like subcellular distribution. The CRISPR/Cas9 approaches have been successful, but homozygosity is yet to be reached, to allow conducting functional assays. Overall, these results shed light on MCL1 3' UTR-APA-derived mRNA isoforms regulation, providing us with opportunities to delve deeper into its subjacent mechanisms and molecular players.

Index

Figures Index:

Tables Index:

Abbreviations:

µL – microliter
3’ UTR – 3’ Untranslated Region
5’ UTR – 5’ Untranslated Region
A1 – BCL-2-related gene A1
Ab – antibody
APA – Alternative polyadenylation
ARE – AU-rich element
AS – alternative splicing
AUE – Auxiliary Upstream Element
BAD – BCL-2 antagonist of cell death
BAK – BCL-2 antagonist killer 1
BAX – BCL-2-associated X protein
Bcl-2 – B cell lymphoma-2
BDNF – Brain-Derived Neurotrophic Factor
BH motif –Bcl-2 Homology motif
BID – BCL-2 interacting domain death agonist
BIK – BCL-2 interacting killer
BIM – BCL-2 interacting mediator of cell death
BMF – BCL-2 modifying factor
Cas – CRISPR-associated
Cdk1 – Cyclin-dependent kinase 1
CDS – coding sequence
ceRNAs – competitive endogenous RNAs
CFIm/ CFIIm – Cleavage Factor Im and/ IIm
Chk1 – checkpoint 1
CLIP – Cross-linking Immunoprecipitation
CPSF – Cleavage and Polyadenylation Specificity Factor
CPSF1 – Cleavage and Polyadenylation Specificity Factor 1
CR – coding region
CRISPR/ Cas9 – Clustered Regularly Interspaced Short Palindromic Repeats
crRNA – CRISPR-RNA
CstF – Cleavage Stimulating Factor
CTD – Carboxy-Terminal Domain
CUGBP2 – CUG triplet repeat RNA-binding protein 2
cyt c – cytochrome c
DN – double-negative
DNA – Deoxyribonucleic acid
DP – double-positive
DSB – double-strand DNA break
DSE – Downstream Sequence Element
E2F1 –E2 Promoter Binding Factor 1
EGFP – Enhanced Green Fluorescent Protein
EtOH – ethanol
FACS – Fluorescence Activated Cell Sorting
FBS – Fetal Bovine Serum
FL – full length
GRE – GU-Rich-Element
gRNA – guide RNA
HDR – homology-directed repair
HIF-1α – hypoxia-inducible factor 1-alpha
HRK – harakiri
HuR – human antigen R
IMM – Inner Mitochondria Membrane
IMP-1 – insulin-like growth factor 2 mRNA binding protein 1
indels – insertions/deletions
INTS9 – Integrator Subunit 9
JAK – Janus kinase
KO – knock-out
LB broth medium – Luria-Bertani broth medium
MAPK – Mitogen-Activated Protein Kinase
MCL1 – Myeloid Cell Leukemia 1
Mcl-1 – myeloid leukemia cell differentiation protein
MCL1ΔN79 – deletion of 79 aa at the Mcl-1 N-terminus
*MCL1*ΔpA1 – *MCL1* with a deleted polyadenylation 1 signal

*MCL1*ΔpA1d – *MCL1* with a deleted polyadenylation 1 signal, using a double sgRNA approach
*MCL1*ΔpA1s – *MCL1* with a deleted polyadenylation 1 signal, using a single sgRNA approach
*MCL1*ΔpA2 – *MCL1* with a deleted polyadenylation 2 signal
*MCL1*ΔpA2 – *MCL1* with a deleted polyadenylation 2 signal, using a double sgRNA approach
MCL1ΔTM – deletion of the transmembrane domain at the Mcl-1 C-terminus
Mcl-1ES – Mcl-1 alternative splicing-derived extra short isoform
Mcl-1L – long Mcl-1 isoform
Mcl-1S – Mcl-1 alternative splicing-derived short isoform
MHC – major histocompatibility complex
miRNA – micro RNA
Mitotracker Red – Mito. Red.
mL – millilitre
MOMP – Mitochondria Outer Membrane Permeabilization
mRNA – messenger RNA
mRNPs – mRNA Nuclear Particles
MS – mass spectrometry
mTORC1 – mammalian target of rapamycin complex 1
NHEJ – non-homologous end joining
nt – nucleotide
O/N – overnight
OMM – Outer Mitochondria Membrane
pA – polyadenylation
pA signal – polyadenylation signal
-pA1 – pEGFP *MCL1* CDS pA1
-pA2 – pEGFP MCL1 CDS pA2
-pA2ΔmiR17 – pEGFP MCL1 CDS pA2ΔmiR17
PABPC – Poly(A)-Binding Protein Cytoplasmic
PABPN – Poly(A)-Binding Protein Nuclear
PAM – protospacer-adjacent motif
PAP – Polyadenylate polymerase
PAS – polyadenylation site
PCNA – Proliferating Cell Nuclear Antigen
PCR – Polymerase Chain Reaction
Pen/ Strep – penicillin/ streptomycin
PEST –proline (P), glutamic acid (E), serine (S), threonine (T)
PI-3K – Akt – phosphatidylinositol 3-Kinase – Serine/ Threonine Kinase
PIP – Protein Interaction Partners
pLenti – sgRNA – lentiCRISPR v2 constructs containing single guide RNA
Poly(A) – Polyadenylate
polyB – Polybrene
PPI – Protein-Protein Interactions
pre-mRNA – precursor mRNA
PTBP1 – polypyrimidine tract binding protein 1
PTM – post-translational modifications
PU.1 – purine-rich box binding
PUMA – p53-upregulated modulator of apoptosis
RBP – RNA Binding Domain – RBD
RBP – RNA Binding Protein
RNA – Ribonucleic acid
RNAPII – RNA polymerase II
ROI – Regions Of Interest
rRNA – ribosomal RNA
RT – room temperature
sgRNA – single-guide RNA
snRNA – small nuclear RNA
SP – single positive
SpCas9 – *Streptococcus pyogenes* Cas9
STAT – Signal Transducers and Activators of Transcription
T lymphocytes/cells – Thymus-dependent lymphocytes/cells
Ta – temperature of primers annealing
TAE – Tris Acetate-EDTA
TALENs – transcription-activator-like effector nucleases
TCRs – T-cell receptors
TEC – thymic epithelial cells
TM – transmembrane domain
TOM – Translocase of the Outer Membrane
tracrRNA – trans-activating crRNA

TTP – tristetraprolin
UPS – Ubiquitin–Proteasome System
USE – Upstream Sequence Element
v/v – one volume
ZFNs – Zinc-Finger Nucleases

Introduction

Gene Expression

Gene expression consists in the conversion of the information encoded by a gene into a functional gene product. The latter can either be a protein – expressed by structural genes through the processes of transcription and translation, or a functional RNA (e.g. ribosomal RNA (rRNA), small nuclear RNA (snRNA), micro RNAs (miRNA) – expressed by non-coding genes[1].

According to the central dogma of molecular biology[2], the information flows from DNA to RNA (transcription) and finally to protein (translation). However, this simple two-step model does not reflect the complexity and interdependency of the processes involved in gene expression, especially in eukaryotic cells, which, for instance, feature transcriptional auto-regulatory feedback loops[3]. Gene expression is subjected to regulation at many levels; starting in the nucleus with a panoply of epigenetic modifications in the chromatin (e.g. chromatin remodeling, histone marks) or in DNA (e.g. base methylation)[4]. Furthermore, gene expression is highly regulated during mRNA synthesis, which in eukaryotes is a dynamic process that occurs in the nucleus and requires a tight connection between transcription, chromatin remodeling and co-transcriptional processing of the precursor (pre-) mRNA[5–7]– an unfinished product of transcription containing both exons and introns. The pre-mRNA processing includes: 5' capping (methylated guanosine – Gm7, addition to the 5'end), splicing (intron excision and exon ligation), 3' end formation and also editing (e.g. base modification[8])[5,7]. The orchestration of these distinct processes is performed by the RNA polymerase II (RNAPII) through its Carboxy-Terminal Domain (CTD), which undergoes differential phosphorylation (among other types of post-translational modifications – PTM) throughout transcription initiation, elongation and termination. The RNAPII CTD allows coupling transcription and pre-mRNA processing in time and space: firstly, processing reactions are modulated by the kinetics/ transcriptional rate of RNAPII, and secondly, CTD acts as a platform to recruit several transcriptional and processing factors, as well as those involved in the packaging of the primary transcript into messenger RNA Nuclear Particles (mRNPs)[5–7].

Polyadenylation

Polyadenylation is a pre-mRNA 3' end processing step, which consists in the addition of a stretch of adenosine monophosphates not encoded by the gene – a polyadenylate [poly(A)] tail, by the poly(A) polymerase (PAP). This occurs after endonucleolytic cleavage of the nascent transcript, determined by the binding of polyadenylation factors[9–13] to the cleavage/ polyadenylation site (PAS). As previously reviewed, some of the factors involved in the formation of the pre-mRNAs 3' ends also associate with transcription initiation and other events of processing[5–7,14–18], like alternative splicing (AS)[19–21]. Furthermore, several polyadenylation factors are involved in setting up appropriate chromatin architecture to aid in efficient transcriptional elongation and termination[22,23].

The pA tail is present at the 3' end of almost all eukaryotic mRNAs (except histone mRNAs[24]) and some long non-coding RNAs, influencing all levels of mRNA metabolism up to maturation, such as: splicing, mRNA export to the cytoplasm, cytoplasmic mRNA stability, translational efficiency and storage[25–30]. The length of the poly(A) tail varies greatly between species (e.g. 250-300 adenines in humans versus 70-80 adenines in yeasts)[25].

There are several *cis*-elements involved in the polyadenylation process, comprising a series of regulatory sequences located mostly in the terminal end of the 3' Untranslated Region (3' UTR), some of which act in concert to determine the choice of a polyadenylation signal – pA signal. The canonical pA signal sequence AAUAAA[11,31], which is the strongest in polyadenylation efficiency, is highly conserved, being found in the vicinity of ~53–58% of human PAS[32–34]. This consensus hexamer acts as a signal for polyadenylation as well as for transcriptional termination[16,17,35], being located 10-30 nucleotides (nt) upstream of the actual PAS – preferably, a CA dinucleotide. Apart from the canonical sequence, the pA signal exists in more than ten weaker variants[10,12,31].

A classical mammalian pA signal associates with another core *cis*-element; a G/U or U-rich Downstream Sequence Element (DSE)[11,36], which may be of supreme importance in the presence of some pA signals of the non-canonical type[37]. Given that database analyses have shown that up to 30% of polyadenylation events do not require the consensus AAUAAA hexamer and ~20% of human PASs are not associated to a U- or GU-rich DSE[32,38,39], auxiliary elements are often present upstream of the PAS – Auxiliary

Upstream Elements (AUEs). Among the latter, the involvement of U-rich Upstream Sequence Elements (USEs) in polyadenylation is the best characterized, being exemplified by Complement C2[40] and COX-2[41].

Furthermore, additional auxiliary *cis*-elements, localized upstream and downstream of the pA signal, are sometimes needed to enhance the polyadenylation of weak pA signals and may help determine the PAS when there are multiple ones[36].

The basal polyadenylation machinery consists of 4 protein complexes : Cleavage and Polyadenylation Specificity Factor (CPSF), Cleavage Stimulating Factor (CstF), Cleavage Factor Im and IIm (CFIm and CFIIm)[42]. CPSF is composed by the following subunits: CPSF-30/CPSF4, CPSF-73/CPSF3, CPSF-100/CPSF2, CPSF-160/CPSF1, Fip1/FIP1L1 and WDC146/WDR33[42,43]. The polyadenylation machinery also includes other proteins, such as PAP, symplekin, PABPN (Poly(A)-Binding Protein Nuclear) and RNAPII[42].

The CPSF complex binds to the CTD through its -160 kDa subunit and to the AAUAAA hexamer through the WDR33 and CPSF30 kDa subunits[44], as was recently shown by structural studies made by Sun and colleagues[44]. It acts in concert with the CstF complex, which binds the GU-rich DSEs through its CstF-64 subunit[10], thus promoting pre-mRNA cleavage between these two *cis*-acting elements. Additionally, the cleavage factors I and II (CFIm and CFIIm) bind upstream of the AAUAAA and are essential for the cleavage step. The PAP is then recruited to this processing complex, resulting in 3' end polyadenylation of the cleaved product. Then, in most eukaryotic species, the poly(A) tails are bound by two important poly(A)-binding proteins (PABPs): PABPN1 in the nucleus and PABPC1 in the cytoplasm, respectively[45].

Regarding transcriptional termination, it is thought that there are multiple processes that can contribute to it, with different types predominating at different genes. For polyadenylated mRNAs, transcription termination often requires the polyadenylation signal, but not the cleavage site, with cleavage-independent and torpedo-style (exonuclease promoted) models of termination being able to occur in parallel[46]. The only crucial factor needed for efficient termination is the necessity of RNAPII to undergo a displacing conformational change and disengage from the DNA[47].

Alternative Polyadenylation

It is now well-established that alternative pre-mRNA processing exerts a very crucial role in modulating gene expression[5–7], mainly because of its ability to generate multiple transcripts harboring different coding sequences. This may be achieved by alternative splicing and/or alternative polyadenylation (APA), when alternative pA signals occur in the intronic or coding regions – internal pA signals (Fig. 1, upper panel), also called CR-APA. These two types of alternative processing give rise to protein isoforms with different functions, which may even be antagonistic when compared to the full-length (FL) isoforms. This is the case of the receptor tyrosine kinase (FL) and its truncated soluble decoy alternative protein isoform with a dominant-negative effect, which comes to lack the anchoring domain upon activation of an intronic pA signal[48].

However, even if alternative transcripts possess the same coding sequence, they can still vary in the length of their 3' UTR, due to APA confined to this region (Fig. 1, lower panel), also named 3' UTR-APA. As insightfully stated by Hughes, untranslated regions are "not just the ends, but the means to an end"[49], because the majority of downstream effects on the transcript's journey are largely dependent on sequences contained in this region. Those effects include: nuclear export, cytoplasmic localization and translation efficiency along with stability[26,50] (Fig. 3, right panel). These processes are mediated by a variable repertoire of RNA Binding Proteins (RBPs), which bind to a range of *cis*-regulatory motifs associated with each of these activities, thus serving as adaptors to a variety of effector proteins[51–53]. Notably, a panoply of novel RBPs (such as some of the heat shock proteins – HSPA1A, HSP90), lacking conventional RNA Binding Domains (RBDs), were recently unveiled[54].

Protein-coding genes can harbor two or more pA signals in their sequences and thus the 3' end formation can occur at different – alternative PAS. Different studies have demonstrated that 70-80% of the mammalian genes contain alternative pA signals, increasing immensely the potential for transcriptome diversity[55,56].

The direct role of RNAPII elongation kinetics in the selection of PASs in the 3' UTR of genes was demonstrated *in vivo* using *D. melanogaster* as a model organism[57,58]. Hypothetically, the slower the elongation rate of RNAPII, the longer the exposure of the proximal PAS to cleavage and polyadenylation factors, favoring its selection by the 3' end

processing machinery over the distal one, according to the "first come, first served" model[12,13]. Hence, the proximal PASs possess an intrinsic advantage over the distal ones. However, in genes subjected to APA regulation, the respective distal PASs may compensate for their unfavorable position relative to the proximal PASs by being associated to *cis*-elements of a more canonical type, including the AAUAAA consensus hexamer and the U/ GU-rich DSE[10,31,59].

Three prime end processing factors relative concentrations influence greatly the APA outcome[13,59]. Thus, the higher the concentration of a particular polyadenylation factor, the greater the probability that it will bind all of its available *cis*-elements, while at lower concentration only the ones presenting a higher affinity will be bound[10]. The first study that confirmed this hypothesis was focused on the APA modulation of the IgM heavy chain pre-mRNA by the CstF-64 levels during B cell differentiation[60]. Afterwards, and in concordance with the previous hypothesis, bioinformatics data showed that the usage of alternative pA signals is indeed biased in different tissues, suggesting a reliance on both *trans*-acting factors and *cis*-regulatory elements[61].

In the light of the given facts, it can be concluded that even small changes in the concentration of any of the factors upon cell transformation and differentiation, will affect the processing of the 3' end and change 3' UTR ratios, thus ultimately impacting gene expression[9,62–64].

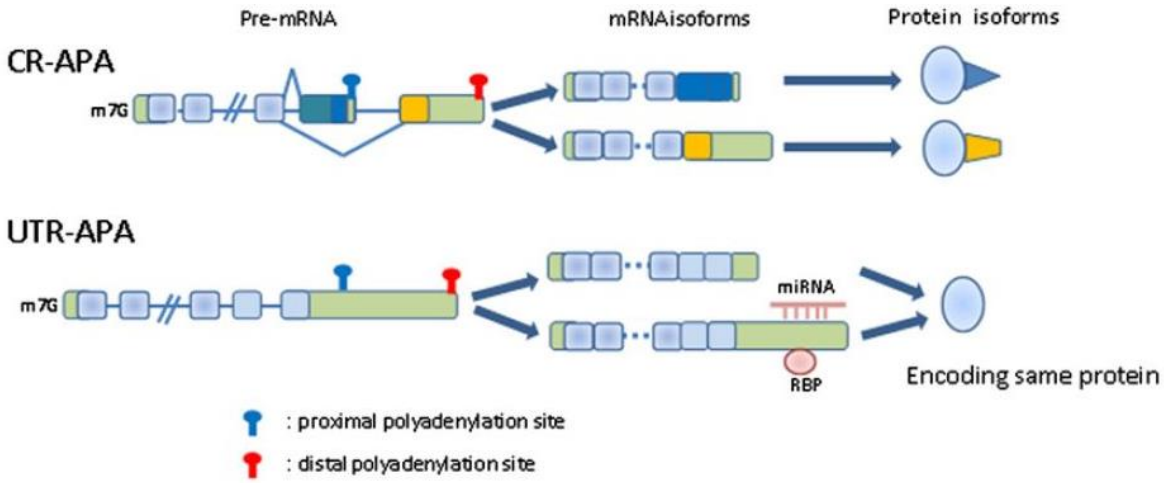

Figure 1 – Alternative Polyadenylation in the coding region (CR) – upper panel, and in the 3' Untranslated Region – lower panel. CR-APA gives rise to different protein isoforms whereas UTR-APA-derived mRNA isoforms encode the same protein. Source: Di Giammartino *et al.*, 2011[10].

The Role Of The 3' UTR

Notably, in humans, the median length of the 3' UTR of genes subjected to APA is of 2,462 nt[55,64], harboring many islands of conserved sequences, with a degree of conservation similar to that of coding regions, and these islands often contain binding sites for miRNAs or RBPs[65,66].

It was shown that 3' UTRs play regulatory roles that are unrelated to the encoded protein, with them being able to act as long noncoding RNAs (e.g. the 3' UTR of the VegT mRNA regulates anchoring of vegetal pole–associated transcripts in *Xenopus* oocytes[67]).

On the other hand, the expression of 3' UTRs can be independent from the coding regions. Through analysis of CAGE libraries of human, mouse, and fly tissues, it was found that a large number of 3' UTRs contained 5' caps and that their expression was often much higher than that of coding regions, sometimes reaching more than a 20-fold difference[68,69].

The 3' UTR-APA mRNA isoforms may be differentially localized in the cell, thus exerting different cellular functions in compliance with spatiotemporal patterns of protein expression and available molecular interactors. That is the case for the Brain-Derived Neurotrophic Factor (BDNF), with the short isoform being restricted to the soma and the long isoform being targeted to dendrites and involved in spine morphology and synaptic

plasticity, in hippocampal neurons[70]. In addition, our group has demonstrated that the longest *RAC1* 3′ UTR is necessary for its transcript targeting to the neurites, and for the subsequent neurite outgrowth in cortical neurons[71].

Apart from mRNA localization, 3’ UTR can also be responsible for localizing the proteins themselves, through an independent mechanism. As recently discovered by Mayr and colleagues[52], the 3’ UTR can modulate differential protein localization due to the establishment of distinct Protein-Protein Interactions (PPI) on the nascent peptide through specific recruitment of protein partners by the 3’ UTR of APA-derived mRNA isoforms. Such is the case for the CD47 protein, which when produced from the long 3’ UTR APA mRNA isoform, localizes to the membrane (Fig. 2), whereas when produced from the short APA isoform, it localizes mainly to the endoplasmic reticulum[52]. Also, as one isoform establishes a particular PPI at the translational level and the other does not, the domains of the latter are available to other interactions. Thus, different 3’ UTR APA-scenarios can ultimately lead to totally different functional outcomes (e.g. apoptosis induction *versus* survival[52]), even if the proteins bear an identical amino acid sequence.

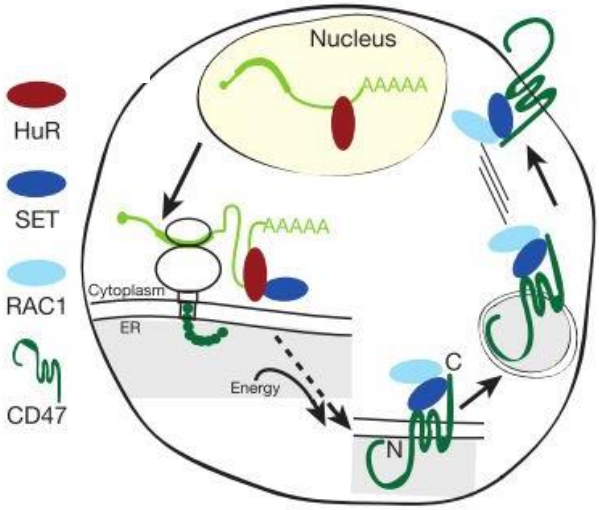

Figure 2 The 3' UTR- dependent protein localization model of CD47.In the nucleus, only the long 3’ UTR APA-derived CD47 isoform recruits HuR, which then recruits SET. SET, by its turn, is transferred onto the nascent protein upon translation and then targets the protein complex to the plasma membrane when associated with RAC1. Source: Mayr and Berkovits (2015)[52].

APA In Physiological Context And In Disease

As the polyadenylation is strongly regulated by *cis*- and *trans*-acting factors, their deregulation can affect gene expression and, ultimately, lead to disease.

Studies have shown that APA is under intense regulation during embryonic development[63,72], presenting a global 3' UTR lengthening pattern[63], which provides a larger regulatory landscape associated with coordinating differentiation and morphogenesis. On the other hand, in cases of response to pathophysiological cues (e.g. HSP70.3 APA in ischemia/ reperfusion injury[73]), in oncogenic transformation[62,74], as well as in cell proliferation[75] , there is often observed a phenomenon of global 3' UTR shortening. In the light of these studies, the shorter APA isoforms were associated with a higher degree of dedifferentiation/proliferation and *vice-versa*[75]. However, one should be mindful that the 3' UTR "shortening" phenomenon, which would be a nuclear co-transcriptional event due to an increased choice of the proximal PAS, may not happen in all the cases. Instead, a differential expression of 3' UTR APA-derived long and short mRNA isoforms may occur at a post-transcriptional level, being modulated mostly by nuclear retention of the longer isoform or by miRNAs or RBPs in the cytoplasm[76].

The consensus explanation for the relationship of shorter APA isoforms with cell proliferation was formulated on the basis that shorter 3' UTRs confer greater stability (which depends on the ratio of synthesis vs decay rates) to the transcripts, partly due to the lack of specific sequences within the truncated 3' UTR, targeted by RBPs and miRNAs that act in a repressive way on mRNA expression[75]. Thus, shorter isoforms were associated with a more efficient protein production[62,75], as was recently demonstrated by our group for the short *CD5* APA-derived isoform, upregulated upon human T-cell activation[77]. However, genome-wide studies revealed that global 3' UTR shortening upon the shift in PAS usage is not always accompanied by increased mRNA stability[78] and/ or protein production[78,79]. These observations highlight the importance of analyzing individual genes in the proper cellular context, to assess the functional relevance of APA isoforms. For instance, the *polo* gene in *Drosophila* is a case in which the efficient translation of the longer isoform occurs, with the deletion of the distal pA signal *in vivo* resulting in the production of only short isoforms and thus leading to aberrant patterns associated with *Drosophila* morphogenesis[57] due to the deficit of Polo protein.

Abnormalities in the 3' end processing mechanism are associated with various hematological, oncological, immunological and neurological diseases, thoroughly reviewed by Curinha and colleagues in 2014[80], which highlights the importance of understanding the APA mechanism and its modulation by *cis*- and *trans*-factors.

For instance, in α and β-thalassemia, abnormalities in polyadenylation and transcriptional termination occur due to point mutations in the proximal canonical pA signal (AAUAAA →AAUAAG for α- and AAUAAA → AACAAA for β- thalassemia), which leads to 3' UTR lengthening, insufficient transcript and hence, low α/ β-globin synthesis[81,82]. In trombophilia, on the other hand, there is a point mutation, which transforms a normal weak PAS variant into a strong one (CG → CA). This leads to a more efficient transcriptional termination due to an increased usage of the PAS and hence to thrombin mRNA and protein overexpression, with deleterious effects[13,83].

In cancer, gene upregulation is often observed, which can be due to regulatory escape caused by 3' UTR shortening upon increased usage of the proximal PAS [e.g. proto- oncogene insulin-like growth factor 2 mRNA binding protein 1 (IMP-1)[84,85]]. Importantly, as recently found in cancer, the proximal PAS shift in oncogenes often disrupts the whole competitive endogenous RNAs (ceRNAs) network, transcripts which act as sponges for miRNAs of their mRNA counterparts, because the oncogenes share binding sites for these miRNAs with tumor suppressor genes. As a result, not only oncogene overexpression is achieved, but also tumor suppressor genes downregulation, as they become more prone to miRNAs regulation[86].

In Alzheimer, in contrast, there is observed a 3' UTR lengthening phenomenon, due to a more frequent usage of the distal PAS upon expression of the cyclooxygenase 2 gene (COX-2). This results in *COX-2* transcript downregulation, caused by the targeting of regulatory sequences in the long 3' UTR isoform, thus leading to disease onset[87].

In conclusion, APA is continuously being demonstrated to play universal and relevant roles both at physiological and pathological level, with global APA regulation (3´ UTR shortening or lengthening) being involved in the ever-increasing number of biological and pathological processes (Fig. 3).

However, "omics" data originated by applying numerous genome-wide high-throughput technologies existent to date[88] need to be combined with focused candidate

gene approaches, in order to unveil the biological function of given genes and also their regulation mechanisms, which when thrown off balance underlie pathological conditions.

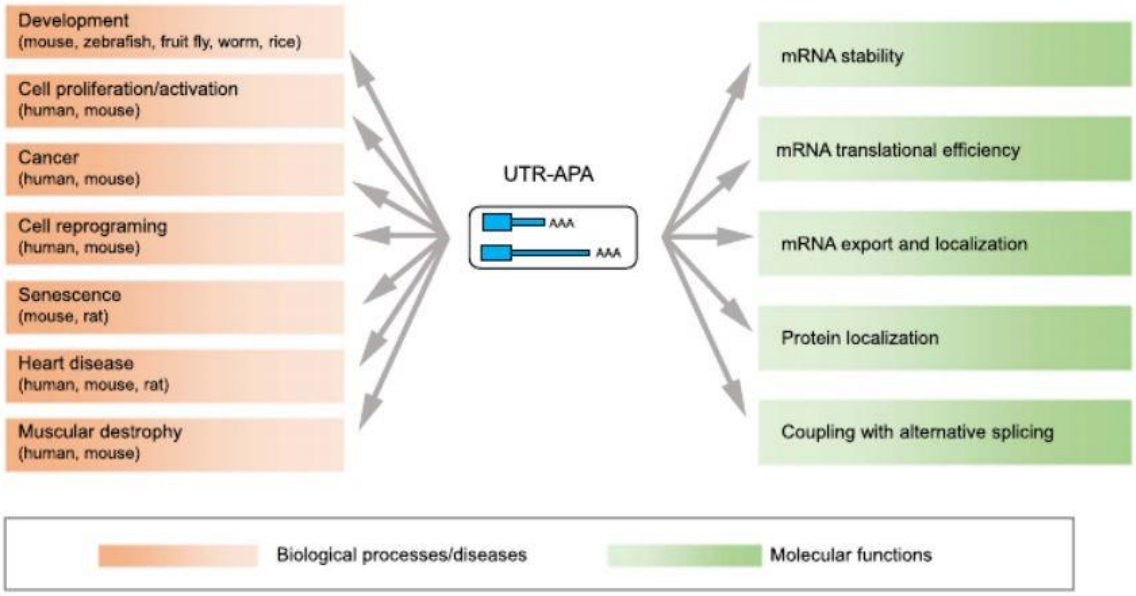

Figure 3 – Functions of Alternative polyadenylation in the 3' Untranslated Region**.** Source: Chen *et al.,* (2017)[88].

Myeloid Cell Leukemia 1 (MCL1)

Myeloid cell leukemia 1 (*MCL1*) gene was first shown to be overexpressed in a human myeloblastic leukemia cell line (ML-1 cells), at an early differentiation stage, by Kozopas and colleagues in 1993[89]. Its protein product induced myeloid leukemia cell differentiation protein (Mcl-1) is associated with poor prognosis in various cancers, such as ovarian carcinoma[90] and certain subtypes of breast cancer, including its aggressive triple-negative form[91], being thus a relevant target for a range of chemotherapeutic agents[92,93].

Mcl-1 is an anti-apoptotic member of the Bcl2 protein family and is unique in that it is crucial for embryonic implantation and early development[94,95], as well as for the survival of multiple cell lineages in the adult[96]. As shown from the analysis of different conditional *knock-out* (KO) mice, Mcl-1 promotes survival in many cell types including; hematopoietic stem cells[97,98], lymphocytes[99–101], neutrophils[102,103], immunoglobulin-secreting plasma cells[104], neural precursor cells[105,106], cardiomyocytes[107] , hepatocytes[108], pancreatic β - cells[109] and thymic epithelial cells (TEC)[110]

Mcl-1 in T-cells Development And Pathophysiology

Thymus-dependent (T) lymphocytes (T cells) are cells of the hematopoietic lineage that develop in the bone marrow from a common lymphoid progenitor with other lymphocytes, such as B lymphocytes and natural killer cells. T cells are distinct from their lymphocyte counterparts by featuring cell surface T-cell receptors (TCRs), which recognize antigen peptide fragments bound to major histocompatibility complex (MHC) molecules on the surface of antigen-presenting cells[111–113]. In this fashion, T-cells are activated, driving antigen-specific immune responses. Thus, the majority of T lymphocytes are master regulators of adaptive cell-mediated immunity, with the exception of γδ T cells[114,115], which present tissue-specific invariant or restricted TCRs, thus displaying innate-like properties.

The name "T cells" arose because, unlike other lymphocyte lineages, even though their precursor cells are generated in the bone marrow, they migrate and colonize the thymus, giving rise to large numbers of thymocytes from which T-cells complete their maturation, through interactions with the thymic stroma[110–113]. In the thymus, developing thymocytes go through a series of distinct phases, marked by changes in the *locus* rearrangement and expression of the TCR (conventional α:β or γ:δ type) gene as well as differential expression of cell-surface proteins; such as the CD3 complex and the co-receptor proteins – CD4 and CD8. Thus, thymocytes firstly differentiate sequentially through four "$CD4^-$, $CD8^-$" double-negative stages (DN1, DN2, DN3, DN4), into double-positive (DP) "$CD4^+$, $CD8^+$" cells. They also undergo negative selection during and after the double-positive stage, which consists in eliminating those cells that respond to self-antigens. Then, DP thymocytes stop expressing either one or the other of the two co-receptor molecules, bifurcating to ultimately differentiate into single positive (SP) – "$CD4^+$,$CD8^-$" or " $CD8^+$,$CD4^-$" thymocytes.

Afterwards, SP thymocytes exit the thymus and enter the peripheral blood. There, they attain functionality following activation and clonal proliferation.[112] Thus, they differentiate from naïve T cells into effector T cells – $CD8^+$ cytotoxic T cells and $CD4^+$ helper T cells. While cytotoxic T cells kill infected cells, helper T cells assist in the activation of macrophages, B cells, and cytotoxic T cells.[112]

In order to determine the role of Mcl-1 in T cell development and function, a variety of studies have been carried out. For example, Opferman and colleagues[97] and Dzhagalov and colleagues[100] showed that T lymphocytes depend on Mcl-1 throughout all of their developmental stages, for survival and maintenance. These authors showed that in the Mcl-1 null mice, DN2 thymocytes presented increased apoptosis whereas DN3 thymocytes showed arrested development, with a decrease in the overall numbers of both DP and SP thymocytes. Moreover, they demonstrated that Mcl-1 has a role in maintaining the local population of mature T lymphocytes, since the deficit in Mcl-1 led to the depletion of T cells from the spleen[97].

Dzhagalov and colleagues[100] confirmed that the *MCL1 knock-out* resulted in the DN3 arrest, as well as in the decrease in the number of SP thymocytes. In addition, they showed that while *in vitro* DP thymocytes depend on Mcl-1 for their survival, *in vivo* its importance is diminished due to the involvement of other anti-apoptotic members[100]. However, later on, Pierson and colleagues[116] showed *in vivo* that Mcl-1 was critical for T regulatory cell survival, with its loss causing fatal autoimmunity, whereas anti-apoptotic proteins Bcl-xL and Bcl-2 were not as important. However, more recently, the relevance for T-cells survival was once again shown to rely on the cumulative participation and relative levels of various anti-apoptotic proteins rather than just one, both *in vitro* and *in vivo*[117].

Regarding the Mcl-1 role in activated T cells, Dzhagalov and colleagues[100] demonstrated that in T-cells purified from mice with *MCL1* tamoxifen-inducible deletion system, activated T lymphocytes had undergone apoptosis upon lack of Mcl-1 expression, being thus clearly reliant on this anti-apoptotic protein. In addition, when T cells were stimulated with an antibody (Ab) binding the CD3 molecule present on T-cells surface (anti-CD3 Ab), a high Mcl-1 up-regulation was observed[100]. Consequently, it was concluded that higher levels of Mcl-1 were needed in T-cells activation, with Mcl-1 being important both for the survival and function of activated T cells[100].

MCL1 Characterization and Regulation

The human *MCL1* gene *locus* resides on chromosome 1q21 and comprises 3 exons[118] that encode the anti-apoptotic protein Mcl-1, which is a member of the B cell lymphoma-2 (Bcl-2) family[118–120]. *MCL1* promoter harbors various elements, such as sis-

inducible element (SIE) – at position -87, and cAMP-response element 2 (CRE-2) – at -70[121,122], which participate in the transcriptional activation of *MCL1*. The latter is subjected to control by several proteins, among which feature growth factors (e.g. epidermal growth factor – EGF[123], vascular endothelial growth factor – VEGF[124]) and a variety of interleukins, namely; IL-3[121], IL-5[125], IL-6[126,127] and IL-7[99]. These cytokines trigger both the Mitogen-Activated Protein Kinase (MAPK) cascade and the phosphatidylinositol 3-Kinase – Serine/ Threonine Kinase (PI-3K – Akt) signaling pathway, leading to the activation of Janus kinases (JAK1, JAK2) and signal transducers and activators of transcription (e.g. STAT3, STAT5), thus ultimately inducing *MCL1* expression and inhibition of apoptosis[128].

In addition, other transcription factors, such as the purine-rich box binding PU.1[129] and the hypoxia-inducible factor 1-alpha (HIF-1α)[130] – on one hand, and the E2 Promoter Binding Factor 1 E2F1[131] – on the other hand, were also described to up-regulate and down-regulate *MCL1* transcription, respectively, by binding to its promoter.

At a co-transcriptional level, *MCL1* is subjected to AS, giving rise to three AS-derived mRNA isoforms,[118,119,132] (Fig. 4). The longest, and the predominantly expressed, AS isoform (Mcl-1L) is anti-apoptotic (Fig. 4), enhancing cell survival through inhibition of apoptosis[118,119,132]. It codifies for one N-terminal PEST domain– peptide sequence that is rich in proline (P), glutamic acid (E), serine (S), and threonine (T), four Bcl-2 Homology (BH) motifs, and one C-terminal transmembrane (TM) domain[119].

However, through skipping of exon 2 or through skipping of a portion of exon 1, two shortened AS mRNA isoforms are generated; Mcl-1S[133] and Mcl-1ES[134] (Fig. 4), respectively. MCL1S lacks the C-terminal BH2 motif and the TM domain, while Mcl-1ES lacks the N-terminal PEST domain and the BH4 motif[133,134]. Contrary to the full-length anti-apoptotic isoform (MCL1L), these two AS-derived isoforms promote apoptosis and induce cell death by, on one hand, sequestering the Mcl-1 anti-apoptotic isoform (dominant negative effect) and on the other hand, being unable to interact and sequester the pro-apoptotic Bcl-2 family members[119] .

In addition to AS, mRNA stability/ turnover is critical in regulation of gene expression and also for mRNA quality control. Generally, the mRNA decay rate is modulated by RBPs bound to its RNA-stabilizing and RNA-destabilizing *cis*-acting elements, commonly located in the 3' UTR. The 3' UTRs' length of the anti-apoptotic

BCL2 family members is quite variable (1506 to 5278 nt). The BCL2 family members also present significant variation in their respective mRNA half-lives[132].

For *MCL1* mRNA, different RBPs bind to its AU-rich elements (ARE) present in the 3' UTR, both stabilizing (by HuR – human antigen R, in glioma[135]) and destabilizing (by tristetraprolin – TTP, in bacterial pathogen-engaged neutrophils[136]), increasing or decreasing gene expression, respectively. In addition, 3' UTR GU-Rich-Elements (GRE) are reported to enhance *MCL1* transcript stability through binding to CUGBP2/ CELF2 (CUG triplet repeat RNA-binding protein 2)[137], even though it suppresses protein expression, whereas 3' UTR CU-rich elements bound by PTBP1 (polypyrimidine tract binding protein 1) destabilize *MCL1* mRNA[138]. However, besides stability, PTBP1 also impacts *MCL1* transcript localization, given that PTBP1 *knockdown* was described to enhance *MCL1* mRNA enrichment in the cytoplasm relative to the nucleus, thus impacting distribution of *MCL1* mRNA, in prostate and lung cancer cells[138].

MCL1 is also down-regulated in the cytosol by miRNAs, being miR-29b[139] the most well-studied.

Finally, at the level of post-translational regulation, *MCL1* is positively regulated through the PI-3K – Akt -mammalian target of rapamycin complex 1 (mTORC1) pathway[140].

Mcl-1 In The context of Apoptosis And The BCL-2 Protein Family

Apoptosis is a programmed form of cell death and occurs under intrinsic (upon DNA damage) or extrinsic (through engagement of death receptors at the plasma membrane, such as Fas) cellular stress conditions. It largely relies on Mitochondria Outer Membrane Permeabilization (MOMP), with a release of the apoptogenic factor cytochrome c (cyt c) from the mitochondria, which subsequently triggers a caspase activation cascade, and then several downstream processes, such as cell dismantling[141,142].

The mammalian BCL-2 protein family consists of both pro-apoptotic and anti-apoptotic members, which constantly interplay, dictating cell fate – survival or apoptosis. The pro-apoptotic members split into 'effectors', which mediate MOMP [e.g BCL-2 antagonist killer 1 (BAK), BCL-2-associated X protein (BAX)]; direct activators, which engage and activate the pro-apoptotic effectors, [e.g. BCL-2 interacting domain death

agonist (BID), BCL-2 interacting mediator of cell death (BIM), p53-upregulated modulator of apoptosis (PUMA)], and derepressors/ sensitizers, which sequester and inhibit anti-apoptotic proteins [BCL-2 antagonist of cell death (BAD), BCL-2 interacting killer (BIK), BCL-2 modifying factor (BMF), harakiri (HRK) and Noxa][120,143]. The anti-apoptotic family members [BCL-2, BCL-xL, BCLw, Mcl-1, and BCL-2-related gene A1 (A1)], on the other hand, inhibit MOMP, by sequestering pro-apoptotic direct activators and effectors, thus antagonizing the direct activator–effector MOMP axes (Fig. 4).

Homology analyses revealed that effector and anti-apoptotic BCL-2 proteins share a high structural homology, namely the four Bcl-2 Homology domains (BH1–4), whose globular folds form a hydrophobic cleft termed the BCL-2 core[144]. However, except BID[145], both direct activators and derepressors/ sensitizers are BH3-only proteins, lacking these globular folds and being intrinsically disordered[146]. Furthermore, many Bcl-2 family members contain a transmembrane (TM) domain at the C-terminus, which allows anchoring to the Mitochondrial Outer Membrane, where they exert their function (e.g. Mcl-1, Bak).

The Mcl-1 protein distinguishes from the other Bcl-2 family anti-apoptotic members because of its short half-life, due to its extensive post-translational regulation. It is a relatively large protein, unlike other anti-apoptotic Bcl-2 members, comprising 350 amino acids (Fig. 4), with 170–300 residues sharing structural/ functional homology to both Bcl-2 and Bcl-XL140. Mcl-1 length is mainly due to its extensive unstructured N-terminal region, which is not present in other Bcl-2 family members and is involved in various post-translational modifications of Mcl-1. Those modulate Mcl-1 turnover and activity, providing the protein with the ability to quickly and reversibly respond to environmental cues, and thus switching cell fate from survival to apoptosis.

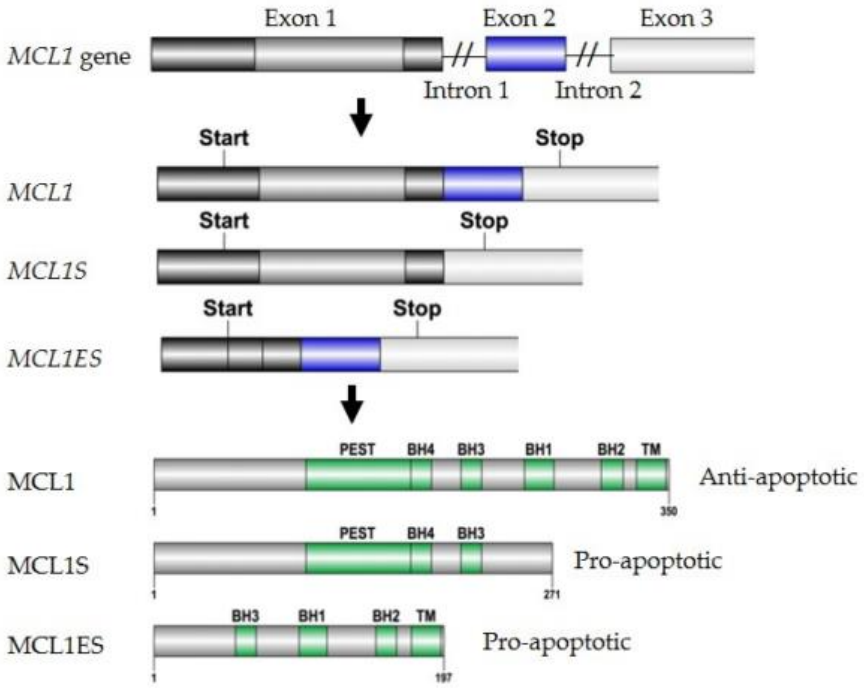

Figure 4 – From top to bottom: MCL1 gene, transcript AS-derived isoforms and protein isoforms. Source: Cui and Placzek (2018)[132].

According to the widely accepted unified model of mammalian Bcl-2 protein family interactions at the mitochondria[120,142,147], anti-apoptotic Mcl-1 blocks MOMP by sequestering the pro-apoptotic effector pore-complex-forming Bak, through binding its BH3 domain. It may also inhibit direct activator BH3-only proteins; Bim, Puma and tBid, depending on cell stress (Fig. 5). On the other hand, Mcl-1 is antagonized by sensitizer BH3-only proteins such as Noxa, which displaces Mcl-1 from Bak and Bax (Fig. 5).[120,147]

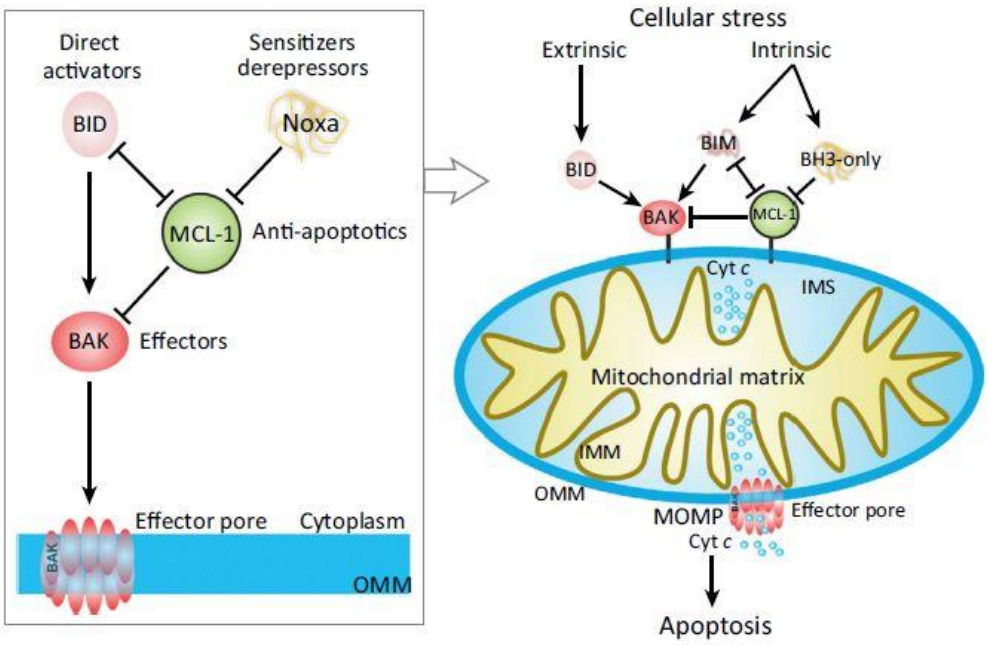

Figure 5 – Bcl-2 family members function in apoptosis. Legend: IMM: Inner Mitochondrial Membrane; OMM: Mitochondrial Outer Membrane; MOMP: Mitochondria Outer Membrane Permeabilization. Source: Moldoveanu *et al*., (2014)[120]

Mcl-1 Localization And Non-Apoptotic Function

The PEST region of Mcl-1, whose reversible phosphorylation regulates this protein stability, also regulates its subcellular distribution, as shown by mutational studies of the residue Ser162, located within this region[148]. Mutation of Ser162 to the phospho-null residue Alanine resulted in predominant nuclear localization, with some cytoplasmic but no mitochondrial localization[148], the latter shown as prevalent by various studies, being where Mcl-1 exerts its anti-apoptotic function when anchored to the OMM[148–150]. This Ser162 mutant Mcl-1 protein isoform also showed a significant decrease in stability and its ability to inhibit Bak-induced apoptosis[148].

Both the deletion of the transmembrane domain at the Mcl-1 C-terminus (MCL1ΔTM) – needed to anchor Mcl-1 to the OMM, and of 79 aa in the N-terminus (MCL1ΔN79) – bearing a mitochondrial targeting signal, were shown to disrupt Mcl-1 mitochondrial localization[149,150]. However, although MCL1ΔTM did not co-localize with the mitochondrial Translocase of the Outer Membrane 20 (TOM20) transport receptor at the mitochondria, MCL1ΔN79 co-localized with TOM20 in 15% of the cells. Curiously, after the inhibition of protein synthesis, the number of cells with mitochondrial

MCL1ΔN79 increased over time, suggesting that the cytosolic localization of MCL1ΔN79 was the consequence of a slower rate of import[150].

Also regarding Mcl-1 targeting to mitochondria, a BH3-only pro-apoptotic protein Noxa/ PMAIP was shown to enhance it in HeLa cells, by establishing PPI with Mcl-1 in the cytosol through the BH3 domain, and hence leading to protein complex co-localization to the mitochondria[151]. Moreover, Noxa was shown to lead to Mcl-1 phosphorylation and subsequent ubiquitination upon entry to the mitochondria, thus enhancing Mcl-1 ubiquitin–proteasome system (UPS)-mediated degradation.

However, Mcl-1 also carries on other functions at the mitochondria, such as regulation of metabolism at the Inner Mitochondrial Membrane (IMM)[152,153], shaping of mitochondria structure and dynamics (fusion/ fission)[153–155] and mitophagy[154].

In addition to its diverse roles at the mitochondria organelle, Mcl-1 was also shown to function in the nucleus. In particular, it regulates cell-cycle progression by engaging Proliferating Cell Nuclear Antigen (PCNA)[156] and Cyclin-dependent kinase 1 (Cdk1)[157] in the nucleus, through its proteolytic protein isoform[157]. Furthermore, Mcl-1 was shown to be specifically and timely accumulated in the nucleus in response to DNA damage, through interaction with the radiation-inducible immediate-early gene IEX-1, while in the absence of the latter, Mcl-1 returned to the cytosol. Both proteins were shown to cooperate to foster checkpoint 1 (Chk1) activation and G2 checkpoint arrest upon DNA damage, with loss of either protein leading to genomic instability and high sensitivity to genotoxic stress[158]. Finally, also in the nucleus, Mcl-1 was demonstrated to regulate the mesenchymal-epithelial transition when associated to STAT3[159], which is needed upon embryonic implantation.

In conclusion, it is evident that Mcl-1 has diverse functions related to cell survival, which is in agreement with the observation of its relevance in so many different cell types and tissues.

CRISPR/ Cas9 Genome Editing System Overview

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and Cas (CRISPR-associated) proteins have been detected in bacteria and archaea for years[160], and discovered only a decade ago to function as a RNA-based adaptive immune system in

prokaryotes, targeting viral genomes for degradation in a sequence-specific manner by the means of a RNA-guided DNA endonuclease[161].

Later on, the most studied type II bacterial-only CRISPR/Cas9 system has been made into the most efficient and powerful tool for genomic engineering to date[162]. It is the simplest CRISPR/Cas9 system, due to concentrating all of its endonuclease activity in a single multidomain protein – Cas9[163], (Fig, 6a).

Prior to the advent of the CRISPR/Cas9 technology, researchers have used naturally occurring[164] and engineered meganucleases[165,166] and engineered nucleases, such as zinc-finger nucleases (ZFNs)[167–169] and transcription-activator-like effector nucleases (TALENs)[170–172], the latter being based on arrays of naturally occurring DNA-binding domains fused to a non-specific DNA-cleaving nuclease. Despite high specificity[172], the design of engineered nucleases was complicated, with a new one being required for each different target *locus*[173]. Thus, CRISPR/Cas9 has come to revolutionize the field of genome editing with its easy programmability[174,175], having made multiplexed gene targeting much more prompt and easy[176,177].

So far, the CRISPR genome editing system has been broadly applied in a variety of organisms, such as mice[107,176], zebrafish[178] and roundworm[179], as well as in different cell types, such as various human cancer cell lines and human pluripotent stem cells[91,180–182], enabling a rapid discovery of new gene functions, developing new cell and animal models of diseases, and making substantial biomedical progresses[183].

In the CRISPR/Cas9 system, the Cas9 endonuclease protein acts in tandem with a ‘‘guide RNA’’ molecule, which localizes it to a specified *locus* in the genome, with a regular guide RNA: gDNA Watson-Crick base pairing following and the Cas9 promoting a double-strand DNA break (DSB) at the target DNA sequence[162,175]

Initially two different RNA molecules, the CRISPR-RNA (crRNA) and the trans-activating crRNA (tracrRNA), were combined to form a functional guide RNA : Cas9 complex[184]. Later on it was discovered that by fusing the crRNA and the tracrRNA into a single-chain chimeric single-guide RNA (sgRNA), the complex was still functional and operated at high frequencies[180]. Thus, the CRISPR/Cas9 system applicability was simplified, requiring only one protein and one small RNA molecule to achieve RNA-directed DNA cleavage[180] .

In response to DSBs induced by Cas9, cellular DNA repair pathways are activated, which mostly lead to random insertions or deletions (indels) occurring at the site of DNA cleavage, repaired predominantly through the non-homologous end joining (NHEJ) enzymatic repair pathway[185,186]. However, the DNA sequences surrounding the cleavage site can be replaced by a homologous sequence through a less frequent homology-directed repair (HDR) mechanism, provided the former exists in the nucleus (G2 and S phase)[185] or if a template donor DNA sequence is introduced. However, given that between the two competing processes, the NHEJ prevails upon the resolution of DSBs, indels are generally a more frequent outcome[162,187–189] (Fig. 6b).

Importantly, the target DNA sequence (the protospacer) must satisfy two basic conditions: be complementary to the guide RNA and present a ‘‘protospacer-adjacent motif’’ (PAM) – a short DNA sequence, usually a trinucleotide, to ensure compatibility with a particular Cas9 nuclease in use[190].

There are various naturally occurring CRISPR nucleases, which have been used to edit mammalian genomes, providing a great targeting flexibility[175]. Each one varies in size, PAM requirement, and location of the introduced DSB within the protospacer, the most commonly used variant being the 1,368-residue long Cas9 protein from *Streptococcus pyogenes* (SpCas9), with a “NGG” PAM (Fig 6b). This relatively simple motif of SpCas9 occurs on average every 8–12 bp in the human genome[191].

The CRISPR/Cas9 technology application is now well established in the coding region, predominantly to create stable KO organisms and cell lines[175,177,179,190,192–194].

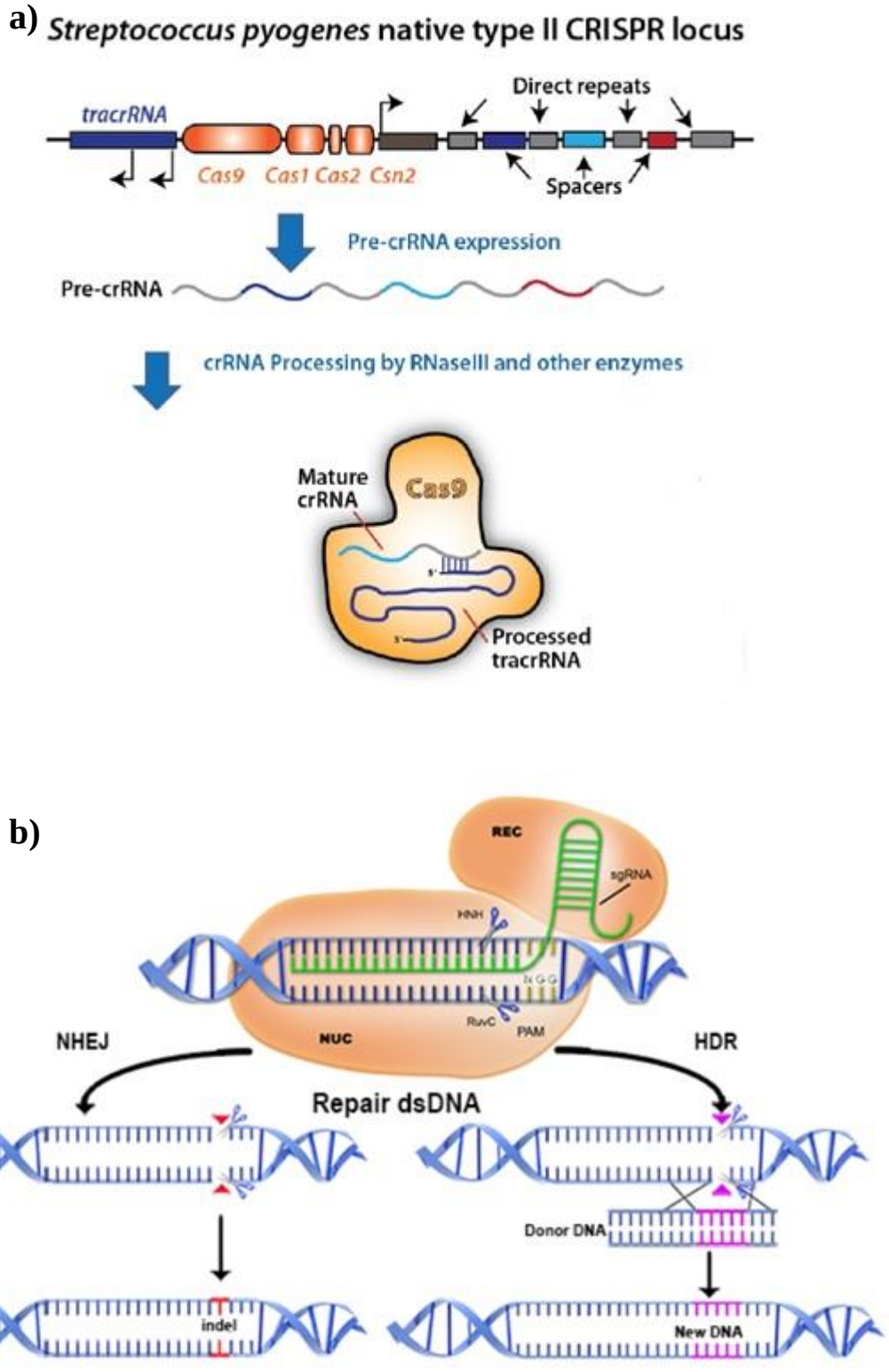

Figure 6 – CRISPR/Cas9 mediated genome editing. **a),** the figure features the biogenesis and assembly of the CRISPR RNA (crRNA) and the *trans*-activating crRNA (tracrRNA), expressed from the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) type II locus of *Streptococcus pyogeenes*, and the introduction of a double strand break (DSB) in the genomic sequence, upon single guide RNA (sgRNA) pairing with its target, containing a protospacer adjacent motif (PAM) trinucleotide**. b),** there is represented a

resolution of a DSB by the means of a Non Homologous End Joining mechanism (left, or through Homology Directed Repair, with a repair template being provided (right). Figure 6a adapted from: PrecisionX™ Cas9 SmartNuclease System Cat. # CAS8/9xxA-1, (2013)[195]. Figure 6b: adapted from Maclean., (2016), *web* page[196]

Background and Aims

Previous work of the lab[197], has shown that the *MCL1* gene harbored four canonical pAs in its extensive (~2,8 kb long) 3' UTR, even though only the two most at the 3' end were used in human T cells (Fig. 7). Those pAs give rise to two APA-derived mRNA isoforms varying only in the length of their 3' UTRs, and whose individual functions are yet to be discovered. The short APA-derived isoform – pA1, has a 3' UTR of 1418 nt long, while the long isoform – pA2 has a 3'UTR of 2828 nt long (Fig. 7). Whilst the pA2 mRNA isoform is more expressed, the pA1 isoform was the one that produced more Mcl-1 protein. In addition, it was shown that the pA2 isoform was down regulated by miR-17 in the cytoplasm[197].

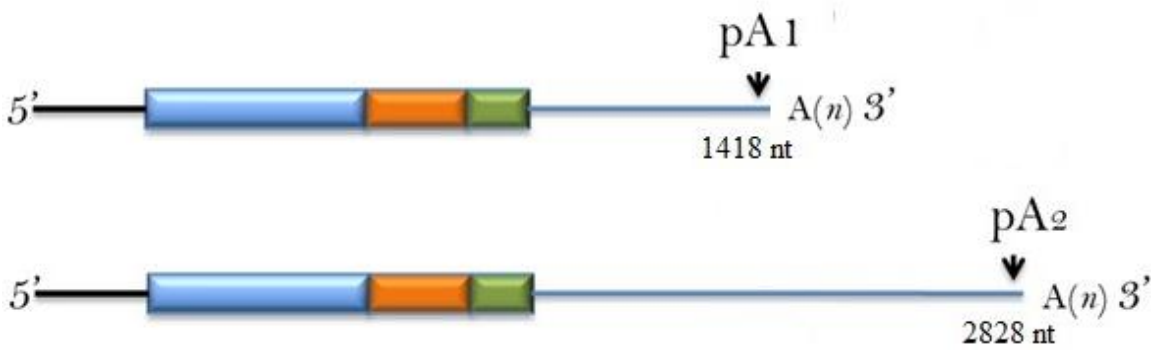

Figure 7 – Scheme of the *MCL1* APA-derived -mRNA isoforms**.** The scheme features the 5' UTR, exons (boxes), and the two 3' UTR lengths with the pA1 **(top)** and the pA2 **(bottom)** polyadenylation signals, plus a polyadenylate tail (A_n) at the end.

On the background of the previous studies, the main aim of this thesis consisted in continuing unraveling the *MCL1* regulatory landscape, by delving deeper into the physiological function of the two APA-derived mRNA isoforms. With that purpose, the first objective of this work consisted in assessing the influence of the two *MCL1* APA isoforms on the localization of Mcl-1 protein. The second objective was to study each isoform separately from one another, by providing a clean, unbiased and physiological

approach to assess pA1 versus pA2 isoforms expression and function, by the means of genome engineering. Finally, it was our aim to assess whether the *knock-out* of two important genes involved in RNA processing: INTS9 and CPSF1, will impact Mcl-1 polyadenylation process and/ or cell viability. The CPSF1 is a basic polyadenylation factor and the INTS9, on the other hand, is part of the Integrator complex, which was revealed to play a termination function at polyadenylated transcripts[198]. In addition, INTS9 *knockdown* may impact the expression of the *MCL1* pA2 isoform, as was suggested by our group preliminary *knockdown* experiments.

Specific aims of this thesis were to:

a) Assess the influence of the two *MCL1* APA isoforms on the localization of Mcl-1 protein;
b) Optimize a method for polyadenylation signal targeting by CRISPR/Cas9 in difficult to transfect cells (E6.1) to generate *MCL1* lacking either the proximal (*MCL1*ΔpA1) or the distal (*MCL1*ΔpA2) pA signals.
c) Obtain CRISPR/Cas9 stable cells lines for CPSF1 and INTS9 *knock-outs*,

Materials and Methods

Cell Culture

Jurkat E6.1 suspension cell line was grown and maintained in complete RPMI 1640 medium, supplemented with GlutaMAX and phenol red, with 10% of Fetal Bovine Serum (FBS) and 1% of a penicillin/ streptomycin (Pen/ Strep) antibiotic solution. Cells were maintained in culture at 37 ºC with 5% of CO_2, in a humidified incubator, being split every 3-4 days (cells passage). At each passage, cells were counted and re-suspended at a 1 x 10^6 cells/mL concentration in fresh complete RPMI medium.

HeLa and HEK 293T adherent cell lines were grown and maintained in complete DMEM medium, with 10% of FBS, and 1% of a Pen/ Strep antibiotic solution. These were maintained in culture at 37 ºC with 5% of CO_2, in a humidified incubator. Every 3-4 days, cells were washed with 1x PBS, trypsinized with TRYple reagent, re-suspended in fresh medium and then passed into a new cell culture flask in a 1:10 dilution, in order to maintain subconfluency (<95%).

All of the aforementioned reagents were provided by Gibco® (Life technologies).

Total RNA Extraction

Jurkat E6.1 cells were harvested [10 million (M)] by centrifuging them in a centrifuge tube at 300 g for 5 min. Thereafter, the supernatant was discarded, the pellet was washed twice with 1,5 mL of 1x PBS, centrifuging at 300 g for 5 min after each washing step. Then, the supernatant was removed, and 1 mL of TRIzol reagent was added to the pellet, homogenizing the solution completely by pipetting up-and-down with a 20 gos (G) needle (~15-17 times). Afterwards, chloroform was briefly added to the lysed sample in TRIzol reagent, in a 1:5 (v/v) ratio. The sample was shaken vigorously in order to mix the contents, and incubated for 3 min, at room temperature (RT). Afterwards, the sample was centrifuged for 15 min at 12000 g, at 4ºC, to promote the separation of the organic and aqueous phases. The upper aqueous phase containing total RNA was transferred into a new centrifuge tube on ice. Thereafter, 1 μL of glycogen (15 mg/mL, Life Technologies) and

one volume (v/v) of isopropanol was added to the total RNA sample to promote RNA precipitation. The sample was vortexed and incubated at -80ºC for at least two hours, to precipitate RNA. After thawing the sample at RT, it was centrifuged for 20 min at 12000 g, at 4ºC. The resulting RNA pellet was washed twice with 500 µL of ice cold 75% ethanol (EtOH) and centrifuged for 10 min at 12000 g, at 4ºC, after each washing step. The pellet was air dried at RT and then re-suspended in 10-30 µL of nuclease-free water (Thermo Scientific™), depending on the pellet size. The sample was then stored at -80ºC, until being used for cDNA synthesis.

cDNA Synthesis

cDNA was synthesized from 1 µg of Jurkat E6.1 total RNA using the SuperScript™ IV (SSIV) Reverse Transcriptase enzyme, according to the manufacturer's protocol (Invitrogen, Life Technologies), subjected to minor changes. RNA samples were diluted into a 14 µL mix, containing nuclease-free water (HyClone™ HyPure Molecular Biology Grade Water, GE Healthcare), 1 µL of dNTPs (10 mM) and 1 µL of random hexamers (50 µM). After a short spin, the mixtures were incubated at 65ºC for 5 min in order to denature RNA, followed by 5 min incubation on ice. Two types of a reverse transcription (RT) master mix were prepared: one with the SSIV enzyme (RT^+) and one without the enzyme (RT^-), a control to screen for genomic DNA contaminations. For each individual sample, the RT^+ mixture's contained 4 µL of a 5X SSIV cDNA synthesis buffer, 1 µL of DTT (0,1 M), 0,5 µL of RiboLock RNase Inhibitor (40 U/µL, Thermo Scientific™) and 0,5 µL of SSIV RT enzyme (200 U/µL, Invitrogen). The RT^- mixture's final composition was the same, except for SSIV addition, which was replaced by nuclease-free water. Samples were incubated for 10 min at 23ºC, then 10 min at 50ºC and finally 10 min at 80ºC to heat-inactivate the enzyme, using a personal thermocycler (Biometra).

Polymerase Chain Reaction (PCR)

Primers with overhangs – 5' CAC**AGATCT**ATGTTTGGCCTCAA 3' Forward (F) and 5' CAC**GAGCTC**CTATCTTATTAGAT 3' Reverse (R), were used to amplify *MCL1* coding sequence (CDS) (1050 bp) containing restriction digestion sites for *Bgl*II at the 5'

end and for *Sac*I at the 3' end (bold sequences), for cloning into a pEGFP-C1 expression vector (Clontech). The reaction mix for each individual sample contained; nuclease-free water, 2 µL of a cDNA sample (previously synthesized from 1 µg of RNA from Jurkat E6.1), 0,2 µL of GoTaq Flexi® DNA Polymerase (#M829A, Promega – 100 U/mL), 4 µL of 5x GoTaq Flexi® Green buffer, 1,5 µL of $MgCl_2$ (25 mM), 1 µL of dNTP mix (10 mM) and 1 µL of each F and R primer (10 µM), in a total volume of 20 µL. Samples were amplified for 37 cycles at 95ºC for 30 s, 58ºC for 30 s, and 72ºC for 60 s, after an initial denaturation step at 95ºC for 3 min, and followed by a final extension step of 5 min at 72ºC in a personal thermocycler (Biometra).

Agarose Gel Electrophoresis

Three µL of SYBR® safe fluorescent nucleic acid dye were used per each 100 mL of 1x TAE (Tris Acetate-EDTA) running buffer. A GeneRuler™ Ladder Mix (Invitrogen Corporation) was used as a molecular weight marker. After completion of PCR reaction, total volumes of the mixes were loaded on a 1% agarose gel and, after 1 h electrophoresis at 120 V, the bands corresponding to the *MCL1* CDS PCR product were excised and then subjected to the Sephadex DNA extraction protocol (see protocol below).

DNA Extraction from Agarose Bands

DNA agarose bands were cut from the agarose gel and placed at -80ºC for at least 20 min. To favor the DNA detachment from the agarose, the 1,5 mL centrifuge tubes containing the bands were incubated for 10 min at 42ºC. Sephadex G-50 (GE Healthcare) fine resine-coated columns were prepared in advance, with 850 µL of Sephadex slurry being pipetted into Zymo-Spin columns (with their silica part being previously removed), centrifuging for 4 min at 2200 g. The columns were then transferred to 1,5 mL centrifuge tubes. After incubation, the partially liquefied agarose bands were vortexed and transferred to their respective Sephadex-coated columns, which were then centrifuged for 10 min at 4400 g. The Sephadex-coated columns filtered out small DNA fragments and agarose, with the DNA sample being eluted into the flow-through. In order to perform DNA ethanol precipitation, 1 µL of glycogen – 20 mg/mL (Roche Diagnostics), 10/3 of absolute ethanol

and 1/3 of ammonium acetate ($NH_4CH_3CO_2$) – 10 M, were added to the samples' volumes, in centrifuge tubes. Samples were then vortexed and centrifuged for 5 min at 14000 g. The resulting pellets were washed with 1 mL of 70% EtOH and centrifuged for 5 min at 14000 g. The DNA pellets were air dried, re-suspended in 10 µL of nuclease-free water and quantified with Nanodrop.

Plasmid Cloning

Firstly, to generate the pEGFP (Enhanced Green Fluorescent Protein) *MCL1* CDS plasmid, the *MCL1* CDS PCR product amplified from E6.1 cDNA was cloned into the pEGFP-C1 expression vector (Clontech), by performing a sequential digestion using two restriction enzymes on both the insert and vector DNAs. *Bgl*II (NEB™ – 10,000 U/mL) was used to cut at the 5' end and *Sac*I (NEB™ – 20,000 U/mL) was used to cut at the 3' end. The 100 µL restriction mixes for *Bgl*II contained: sterile water, 10 µL of 10x NEB™ 3.1 buffer, 3 µg of DNA sample and 1 µL of restriction enzyme. Also, a non-cut control without restriction enzyme was made for the vector. After 2 h digestion at 37ºC, 100 µL of sterile water was added to the samples, followed by the addition of one volume of Phenol/ Chloroform/ Isoamyl Alcohol solution (25:24:1; Sigma). The samples were vortexed and centrifuged for 5 min at 14000 g, and the aqueous phases were recovered and subjected to EtOH precipitation, as described above. DNA pellets were re-suspended in 44 µL of sterile water, with 5 µL of 10x NEB™ 1.1 buffer and 1 µL of *Sac*I enzyme being added. *Sac*I digestion was performed for 2 h, at 37ºC. Then, 1 µL of FastAP thermosensitive alkaline phosphatase (Thermo Scientific™ – 1 U/µL) was added to the vector and incubated for 30 min at 37ºC. Thereafter, the enzymes were heat-inactivated, by incubation at 65ºC for 20 min. Samples were loaded into a 0,8% agarose gel, and the bands corresponding to the inserts and vector were excised to purify DNA through the Sephadex protocol (see section above). Then, a subcloning step was performed, consisting in the addition of *MCL1* 3' UTR sequences downstream of CDS, namely: a short pA1 isoform 3' UTR, a long pA2 isoform 3' UTR or a pA2 isoform 3' UTR with a mutated miR-17 binding site. The two pA2 isoform 3' UTR sequences had the upstream pA1 signal mutated. Thus, to generate the following plasmids: pEGFP *MCL1* CDS pA1, pEGFP *MCL1* CDS pA2 and pEGFP *MCL1* CDS pA2Δmir17, *Sac*I – at the 5' end, and *Sal*I (NEB™ – 100,000 U/mL,

10xNEB™ 3.1 buffer) – at the 3' end, restriction enzymes were used as described above to cut the pEGFP *MCL1* CDS vector and the inserts from already available pEGFP-pA1/ pA2/ Luc-pA2Δmir17 plasmids.

DNA Insert Sticky-End Ligation Into Vector DNA

Firstly, the NEB Biocalculator online software was used to determine the mass (ng) of insert that should be used in a ligation reaction at a given insert to vector DNA molar ratio (3:1 ; 5:1). The lengths of insert (*MCL1* CDS PCR product –1050 bp; pA1 isoform 3' UTR – 1418 bp; pA2 isoform 3' UTR – 2828 bp, and pA2Δmir17 isoform 3'UTR – 2828 bp) and vector DNA (pEGFP-C1 – ~4,7 kb , pEGFP *MCL1* CDS – ~5,8 kb), as well as the set mass of linear vector DNA (100-150 ng) were taken into account. Each ligation mix, besides vector and insert DNA samples, contained sterile water, 1 µL of T4 DNA Ligase (#EL0011, Thermo Scientific™ – 5 U/µL) and 2 µL of 10x T4 DNA Ligase Buffer, in a total reaction volume of 20 µL. Also, two controls were made for each vector-insert combination: Vector + Ligase and Vector – Ligase, in order to provide a background reference at the transformation step, in terms of the number of colonies. After 2 h incubation at 22ºC, the T4 DNA Ligase enzyme was heat inactivated, by incubating at 65ºC for 10 min. All reagents were provided by Thermo Scientific™.

Transformation Of Chemically Competent Bacteria

The ligation solution containing plasmid DNA was added in a 1:10 (v/v) ratio to the 100 µL of Top10 *E.coli* strain chemically competent bacteria, incubating on ice for 15 min. Then, a heat-shock was performed by incubation for exactly 1 min at 42ºC, in a thermostat (Eppendorf). Thereafter, the transformation mixes were once again incubated on ice for 5 min, followed by an addition of 4 volumes of a Luria-Bertani (LB) broth medium. The transformants were incubated in an orbital shaker for 1 h at 37ºC. The transformation mixes were plated out on petri dishes containing LB agar selection medium supplemented with 50 µg/mL Kanamycin or 100 µg/mL Ampicillin, depending on the selection marker encoded in the vector. The volume and concentration of transformed *E. coli* varied with the type of vector being used. For plasmid transformation, 10 µL of bacterial suspension was

plated out directly, whereas for ligation mixes, cells were centrifuged at 400 g for 2 min, with most of the supernatant being discarded and the whole pellet re-suspended in the remaining supernatant being plated out. Petri plates were incubated overnight (O/N) at 37ºC and cell growth was assessed on the following day, with round isolated colonies being selected for colony PCR or inoculation.

Colony PCR

For one sample, the colony PCR reaction mix had a total volume of 10 µL, containing; 5,6 µL of sterile water, 2 µL of 5x GoTaq® Flexi polymerase Green buffer (Promega Corporation), 0,75 µL of $MgCl_2$ solution (25 mM), 0,5 µL of dNTPs mix (10 mM), 0,5 µL of each – forward and reverse, primer (10 µM) and 0,15 µL of GoTaq® Flexi polymerase enzyme (Promega Corporation). Only isolated colonies were picked for the colony PCR reaction. In parallel, same colonies were inoculated into their corresponding wells on a 96-well cell culture plate. To this end, 100 µL of LB broth medium supplemented with antibiotic was pipetted per well in advance, accounting for samples and one negative control. Thus, colonies were sequentially dipped into their respective PCR mixes, using pipette tips, and then into the cell culture plate wells. A clean pipette tip was placed into the negative control PCR tube and well. . Samples were amplified for 35 cycles at 95ºC for 45 s, at Ta (temperature of primers annealing –adjusted) for 45 s, and at 72ºC for a time depending on the amplicon size (~1 min/ 1kb), after an initial cell lysis/ denaturation step at 95ºC for 5 min, and followed by a final extension step of 5 min at 72ºC in a personal thermocycler (Biometra).

Throughout the PCR reaction, the inoculated clones were incubated at 37ºC. After completion, whole PCR reaction volumes were loaded into a 1% agarose gel and run for ~30 min at 120 V. After electrophoretic profile analysis, the alleged positive clones were inoculated in 5 mL of LB broth supplemented with antibiotic.

Plasmid DNA Extraction

The ZR Plasmid Miniprep™ kit (ZYMO Research) protocol subjected to minor changes was applied to extract the plasmid DNA from the transformed competent bacteria.

Firstly, to obtain a 4,5 mL-*E. coli* suspension pellet, the same centrifuge tube was filled thrice at maximum capacity, centrifuging at 16000 g for 30 sec, and discarding the supernatant. Thereafter, 200 µL of P1 Buffer was added to the tube, and the pellet was re-suspended completely by vortexing. Afterwards, 200 µL of P2 Buffer was added to the cell suspension, mixing contents by inverting 4 times and letting the solution incubate for 2 min at RT, to promote cell lysis. P3 buffer (400 µL) was added to the tube, mixing contents by inverting 4-5 times and incubating the lysate for 2 min at RT, to promote neutralization. Thereafter, the sample was centrifuged at 16000 g for 4 min, and the supernatant containing plasmid DNA was carefully transferred to a Zymo-Spin IIN column placed into a collection tube. This Zymo-Spin IIN assembly was centrifuged for 30 sec at 16000 g, the flow-through was discarded, and 200 µL of Endo-Wash Buffer were added to the column. The assembly was centrifuged for 30 sec at 16000 g, and 400 µL of Plasmid Wash Buffer was added, centrifuging the assembly for 1 min at 16000 g. Lastly, the plasmid DNA was eluted in 30 µL of nuclease-free water into a clean centrifuge tube, by adding water directly to the column matrix, incubating for 1 min at RT, and then centrifuging for 1 min at 16000 g. The eluted plasmid solution was stored at -20ºC. All plasmids were sequenced on an ABI 3130XL Automated Sequencer (Applied Biosystems) to confirm that they had been accurately constructed.

HeLa Cells Transfection

For each experiment condition, 120 000 HeLa cells were seeded the day before transfection on a µ-Slide 4 Well Glass Bottom (Ibidi®), in 750 µL of complete DMEM medium and at standard culture conditions, to reach a 60–80% confluency. HeLa cells transfection with EGFP-containing constructs, namely: pEGFP *MCL1* CDS pA1, pEGFP *MCL1* CDS pA2 and pEGFP *MCL1* CDS pA2Δmir17, was carried out the day after seeding using the jetPRIME transfection reagent (Polyplus-transfection®), according to the manufacturer's standard protocol. Briefly, media were removed from the cells, and 500 µL of fresh media without Pen/ Strep was added per well. Afterwards, 0,5 g of each EGFP-containing construct was diluted in 50 µL of jetPRIME buffer, with the solutions being vortexed for 10 sec and then short-spinned. One µL of jetPRIME reagent was added to the previous solutions in a 1:2 plasmid weight per volume of jetPRIME reagent proportion

(w/v). The solutions were vortexed for 10 sec, short-spinned and incubated for 10 min at RT, to form micelles. After that, the transfection mixtures were added dropwise to the cells, being homogenized by rocking the cell culture vessel. Media were replaced by 750 µL of complete DMEM media 4-24 hrs after transfection. HeLa transfection efficiency was assessed 48 h post transfection, through observation of the EGFP fluorescence in a ZOE inverted fluorescence microscope (Bio-Rad).

MitoTracker Red and Hoechst Live Cell Staining

Firstly, the media were removed from the µ-Slide 4 Well Glass Bottom (Ibidi®) wells. Then, for mitochondria staining, cells were incubated in 500 µL of complete DMEM medium supplemented with 50 nM of MitoTracker Red solution (Thermo Scientific™), for 1 h, at 37ºC, in a 5% CO_2 humidified incubator. Thereafter, for nuclei staining, cells were washed twice with 1x PBS and then incubated in 500 µL of complete DMEM medium supplemented with 0,5 µg/mL of Hoechst 33342 (Thermo Scientific™), for 30 min at 37ºC in a 5% CO_2 humidified incubator. Cells were washed twice with 1x PBS and then, 500 µL of a 1% FBS/ HBSS solution was added to the samples.

Confocal Microscopy Live Cell Imaging

Images were acquired in a Sp5 inverted confocal microscope (Leica), using the 63X oil objective with a 1024x1024 pixel resolution. The settings were maintained the same for all of the three constructs (pEGFP MCL1 CDS pA1, pEGFP MCL1 CDS pA2 and pEGFP MCL1 CDS pA2Δmir17), with the following lasers set being used: 405 nm for Hoechst, 488 nm for EGFP and 561 nm for MitoTracker Red. Images were processed and analyzed using the FIJI package for Image J software. To quantify EGFP fluorescence intensity, firstly, the areas and the mean fluorescence intensity values (in arbitrary units – a.u.) for each transfected cell were measured using the ROI (Regions Of Interest) manager tool available in the ImageJ software. Thereafter, those mean fluorescence intensity values were corrected by subtracting the mean value of background fluorescence intensity and the obtained values were then multiplied by the area of each cell, with total fluorescence intensity levels being thus quantified for ~50 cells per each transfection condition (pEGFP

MCL1 CDS pA1/ -pA2 and -pA2ΔmiR17). In order to quantify statistical differences, the total fluorescence intensity values for cells in each of these three conditions were uploaded into the GraphPad Prism 7.04 *software*. Then, an unpaired *t* test with Welch's correction was applied to the pEGFP *MCL1* CDS pA1 vs -pA2 and -pA2 vs -pA2ΔmiR17 datasets.

Guide RNAs Design Strategy

In order to obtain CPSF1 and INTS9 gene knock-outs (KO) as well as *MCL1*ΔpA1 and *MCL1*ΔpA2 Jurkat E6.1 stable cell lines, a CRISPR/Cas9 approach was chosen and guide RNAs (gRNAs) were designed using the "CRISPR Design" online software (available on http://crispr.mit.edu/). The gRNA selection was based on score and on the number of off-targets. Guide RNA oligo sequences, scores and the number of off-targets are listed in table 1. For CPSF1 and INTS9 gene knock-outs, one pair of guides (one upstream and one downstream) was designed, matching the coding sequence. For *MCL1*ΔpA1, a single-guide (*MCL1*ΔpA1s) as well as a double-guide (*MCL1*ΔpA1d) strategy was developed. The former consisted in the double gRNA targeting of the pA1 signal neighboring region (with a predicted cleavage site being only 5 nt apart from the hexamer), while the latter consisted in the targeting of genome sequences upstream and downstream of the pA1 signal, to delete a ~70 nt region including the pA1 signal. For *MCL1*ΔpA2, two different pairs of gRNAs were designed: upstream and downstream of the pA2 signal. These guides created a deletion (1st guide pairs around 200 nt and 2nd guide pairs around 300 nt) including the pA2 signal and its respective neighboring region.

For each gRNA, a forward (F) and a reverse (R) oligonucleotide (oligo) was ordered (table 1). The F oligo contained a 5' end – "CACCG" overhang, whereas the R oligo contained a 5' – "AAAC", together with a 3' end – "C", overhangs. These overhangs were designed with a purpose to provide *Bsm*BI restriction enzyme sites, for cloning into the all-in-one lentiCRISPR v2 (#52961, Addgene) lentiviral vector (see section bellow: gRNAs Cloning Into The All-in-One Vector).

Table 1 – Guide RNAs oligonucleotides list and guides features.

Guide RNA oligo name	Target region	Sequence (5'-3') with overhangs	Score http://crispr.mit.edu/	Strand	Sequence lenght
CPSF1_gRNA1-F	CPSF1	caccgATGTACACACGCCGCGAGTG	96	(+)	25
CPSF1_gRNA1-R		aaacCACTCGCGGCGTGTGTACATc			
CPSF1_gRNA2-F		caccgCCATGCTTGTCTACGGCACG	95	(+)	
CPSF1_gRNA2-R		aaacCGTGCCGTAGACAAGCATGGc			
INTS9_gRNA1-F	INTS9	caccgACGGAACCCACCGTCCAGAT	93	(+)	
INTS9_gRNA1-R		aaacATCTGGACGGTGGGTTCCGTc			
INTS9_gRNA2-F		caccgACCGAGCACACCGGCTTCAC	88	(+)	
INTS9_gRNA2-R		aaacGTGAAGCCGGTGTGCTCGGTc			
pA1single_gRNA-F	MCL1 pA1	caccgATTTGCTTTATTGACATACT	42	(-)	
pA1single_gRNA-R		aaacAGTATGTCAATAAAGCAAATc			
pA1up_gRNA-F		caccgGGCTTGCTTGTTACACACAC	79	(+)	
pA1up_gRNA-R		aaacGTGTGTGTAACAAGCAAGCCc			
pA1down_gRNA-F		caccgCAATGCAAAAACTTGCAACA	60	(-)	
pA1down_gRNA-R		aaacTGTTGCAAGTTTTTGCATTGc			
pA2up1_gRNA-F	MCL1 pA2	caccgGTTCGGGCAAATCCTCCAAA	79	(+)	
pA2up1_gRNA-R		aaacTTTGGAGGATTTGCCCGAACc			
pA2up2_gRNA-F		caccgGCTGACTGGCTACGTAGTTC	92	(+)	
pA2up2_gRNA-R		aaacGAACTACGTAGCCAGTCAGCc			
pA2down1_gRNA-F		caccgTCAAATACAGGGTGTGATAT	74	(+)	
pA2down1_gRNA-R		aaacATATCACACCCTGTATTTGAc			
pA2down2_gRNA-F		caccgTCAGCTGAGCAAATATGTAC	70	(+)	
pA2down2_gRNA-R		aaacGTACATATTTGCTCAGCTGAc			

LentiCRISPR v2 Plasmid Purification

The all-in-one lentiCRISPR v2 (Addgene #52961) lentiviral vector, encoding a humanized *Streptococcus pyogenes* Cas9 (hSpCas9) and the *Bsm*BI restriction site for insertion of a guide RNA sequence, was purified using the HiSpeed® Plasmid Purification Maxi Kit (Qiagen) using the manufacturer's protocol instructions being eluted in 500 µL of nuclease-free water. This plasmid was digested with a *Bsm*BI (NEB™ – 10,000 U/mL), being thereafter purified from the agarose gel using the Sephadex protocol (see section above).

gRNAs Cloning Into The lentiCRISPR v2 All-in-One Vector

The gRNAs' F and R oligos were annealed and then cloned individually into the all-in-one lentiCRISPR v2 lentiviral vector previously digested with *Bsm*BI. Firstly, 1 µL of forward and 1 µL of reverse oligo (100 µM) for each of the 11 gRNAs pairs (table 1), were combined in a 50 µL mix, also containing 5 µL of a T4 DNA Ligase 10x buffer and sterile water. The annealing reactions were put in a thermocycler at 95ºC for 5 min, being finally cooled down to 25ºC, at a 5ºC/min rate. Thereafter, the annealed oligos mixes were diluted at a 1:10 dilution into sterile water. Afterwards, 2 µL of T4 DNA Ligase 10x

buffer, 1 µL of T4 DNA Ligase enzyme, 1 µL of a diluted annealed oligos solution (insert) and 1 µL of the lentiCRISPR v2 vector were combined in a ligation mix, brought up to 20 µL with nuclease-free water, for each of the 11 constructs. The inserts' and the vector's DNAs sticky ends were ligated as described in the above section "DNA Insert Sticky-End Ligation Into Vector DNA". The obtained constructs thus contained the chimeric single guide RNA (tracrRNA fused with gRNA: sgRNA) and the hSpCas9 expression cassettes – pLenti –sgRNA constructs. Thereafter, those were transformed in chemically competent bacteria (see section above), with the trasnformants being plated ot on LB agar plates supplemented with 100 µg/mL ampicillin. A colony PCR was performed using the primers listed below (table 2). Guide RNA sequences with a 5' overhang were used as the forward primers, and a sequence matching the lentiCRISPR v2 backbone was used as a common reverse primer. Thereafter, positive clones were inoculated and the plasmids extracted using the ZR Plasmid Miniprep^{TM} kit (see the "Plasmid Extraction" section above). All obtained plasmids were sequenced as referred in the "Plasmid Extraction" above.

Table 2 – Colony PCR primer list for the pLenti – sgRNA constructs.

Primer Forward (F)	Primer F sequence	Primer Reverse (R)	Primer R sequence
CPSF1_gRNA1-F	caccgATGTACACACGCCGCGAGTG	Lenticrispr-R	TGGCACCGAGTCGGTGCTT
CPSF1_gRNA2-F	caccgCCATGCTTGTCTACGGCACG		
INTS9_gRNA1-F	caccgACGGAACCCACCGTCCAGAT		
INTS9_gRNA2-F	caccgACCGAGCACACCGGCTTCAC		
pA1single_gRNA-F	caccgATTTGCTTTATTGACATACT		
pA1up_gRNA-F	caccgGGCTTGCTTGTTACACACAC		
pA1down_gRNA-F	caccgCAATGCAAAAACTTGCAACA		
pA2up1_gRNA-F	caccgGTTCGGGCAAATCCTCCAAA		
pA2up2_gRNA-F	caccgGCTGACTGGCTACGTAGTTC		
pA2down1_gRNA-F	caccgTCAAATACAGGGTGTGATAT		
pA2down2_gRNA-F	caccgTCAGCTGAGCAAATATGTAC		

Packing 293T Cell Line Transfection

Three hundred thousand HEK 293T cells were seeded in a 6-well cell culture plate for each of the eleven pLenti – sgRNA constructs (see sections above) to be transfected, plus one non-transfected control. Twenty-four hours after cell seeding, cells were

transfected using a standard manufacturer's "jetPRIME® in vitro DNA & siRNA transfection reagent" protocol (Polyplus®). Three vectors were co-transfected: a packaging plasmid – pCMV-dR8.91 (Addgene), an envelope plasmid (pMD4) and a transfer vector: lentiCRISPR v2 constructs with guide RNAs and Cas9. For a total plasmid mass of 2 µg, different mass ratios of the three types of plasmids were used: 850 ng for the lentiCRISPR v2 construct, 750 ng for pCMV-dR8.91 and 400 ng for pMD4. The masses of the plasmids used were based on their size – the bigger ones were transfected in higher proportion.

Lentiviral Transduction

Eight pLenti – sgRNA constructs conditions were planned out, by using eleven sgRNAs (table 1) lentiviral constructs: CPSF1 KO – upstream (up) and downstream (down) guide; INTS9 KO – up + down guide; *MCL1*ΔpA1s – single guide; *MCL1*ΔpA1d (double guide) – up + down guides; *MCL1*ΔpA2d – up1 + down1, up2 + down2, up1 + down2 and up2 + down1 guides. As lentiviral transduction was to be carried out in Jurkat E6.1 cells, DMEM media were replaced by complete RPMI media 4 h after packing 293T cell line transfection. Forty eight hours post-transfection, the lentiviral particles-containing 293T media were collected into 15 mL falcons. Those were centrifuged at 3000 rpm for 5 min at RT, to get rid of any 293T cells and debris, thereafter transferring the virus-containing supernatants into new falcon tubes. For each experiment condition, 1 x 10^6 Jurkat E6.1 cells were re-suspended in 2 mL of virus supernatant and transferred to T25 flasks. When combining lentiviruses, 1 mL of each virus supernatant was used (storing the remainder at -80ºC). Three mL of fresh RPMI complete media was added to each T25, with the media final volume being then supplemented with 8 µg/mL of Polybrene (polyB). Seventy-two hours past, the media were changed, with the cells being re-suspended in fresh RPMI complete media supplemented with 5 µg/mL of puromycin (antibiotic resistance encoded in the lentiCRISPR v2 vector backbone).

Genomic DNA Extraction From Cell Lines

One mL of E6.1 cell suspension from each pLenti – sgRNA constructs (see in the section above) was collected in an eppendorf tube and centrifuged for 5 min at 300 g.

Pellets were washed with 1 mL of 1x PBS, centrifuged, and then re-suspended in 500 µL of 1x STE buffer (100 mM NaCl, 10 mM Tris-Cl, pH 8.0 and 1 mM EDTA). Twenty-five microliters of SDS 20% and 10 µL of proteinase K (20 mg/mL) was added to the samples. Samples were incubated O/N at 56°C, in an orbital shaker. On the next day, 20 µL of 5M NaCl solution was added to the samples. Then, in a fume hood, 550 µL of a Phenol: Chloroform: Isoamyl Alcohol (25:24:1) solution was added. Samples were vortexed and then centrifuged for 5 min at 14000 g at RT. Five hundred and fifty microliters of chloroform was added to the recovered aqueous phases and samples were vortexed and centrifuged for 5 min at 14000 g, at RT. Aqueous phases were recovered once again (~500 µL), and 1 mL of absolute ethanol (~2 volumes), 50 µL of Sodium Acetate 3M pH 5.2 (~1/10 volume) and 1 µL of glycogen (20 mg/mL) was added. Thereafter, samples were incubated for at least 2 h at -80°C, to precipitate DNA. Then, those were centrifuged for 20 min at 4°C at top speed, and the supernatants were discarded. DNA pellets were washed with 1 mL of 70% EtOH, being centrifuged for 10 min at 4°C at top speed. Finally, pellets were air dried and re-suspended in 50 µL of nuclease-free water.

Genomic DNA Extraction From E6.1 Growing in a 96-well Plate

The 96-well plates containing Jurkat E6.1 were centrifuged at 500 g for 5 min to pellet the cells and media were removed. Thereafter, cells were washed once with 150 µL of 1x PBS and 50 µL of Bradley Lysis Buffer (10 mM Tris- HCl (pH 7.5), 10 mM EDTA, 0.5 % SDS and 10 mM NaCl) supplemented with 1 mg/mL of proteinase K was added to the samples' wells. After that, the lid was replaced and the plate was sealed with parafilm, being incubated O/N at 60ºC, in a humidified chamber. After being cooled down to RT, 100 µL of an ice- cold EtOH/NaCl solution (75 mM NaCl in 100% EtOH) was added to the samples to precipitate DNA, mixing well the contents. After incubation at RT for 30 min, the plate was centrifuged for 20 min at 3000 rpm, the supernatant was removed and the pellets were rinsed twice with 150 µL of ice-cold 70% EtOH, centrifuging the plate for 10 min, at 3000 rpm. Afterwards, the supernatants were discarded and the DNA pellets were allowed to air- dry and ressuspended in 30 µL of warm nuclease-free water. The 96-well plate was incubated at 56 ºC for 10 min to promote DNA pellets resuspension.

Samples were quantified using a Nanodrop™ 1000 Spectrophotometer (Thermo Scientific™).

CRISPR/Cas9-subjected Cells Genotyping

Genotyping was performed on each of the 8 pLenti – sgRNA constructs conditions on bulk E6.1 cell populations with 10 days of puromycin selection, and on clonal populations derived from single-cell sorting into 96-well cell culture plates (see section below). A PCR reaction mix for an individual sample contained: 4 µL of 5x GoTaq Flexi® Green buffer, 1,5 µL of $MgCl_2$ (25 mM), 1 µL of a dNTPs mix (10 mM), 1 µL of each primer (10 µM) – forward and reverse (table 3), 0,2 µL of GoTaq Flexi® polymerase (#M829A, Promega – 100 U/mL) and 1 µL of gDNA sample (~50 ng/µL). The mix's volume was completed to 20 µL by adding nuclease-free water.

Table 3 – Primers used for genotyping.

Target gene	Primer	Sequence	Amplicon size w/o deletion (bp)
CPSF1	PCR1_2-F	TTCTGTGGCCCCAAGTGTTT	250
	PCR1_2-R	CACCACAGCCTGCCTCAG	
INTS9	PCR3_4-F	ATTCCCTCCCTTTCTTTTCAGACG	267
	PCR3_4-R	TCAGAAGGAACTAACCACCAGC	
MCL1 pA1	MCL1_pA1.2-F	GATGGCTTGGAAAAGCAGGC	254
	MCL1pA1-R-XL	ATC TGT AGA GGG AGC AGA AC	
MCL1 pA2	MCL1_pA2-F2	GGAGGAGGAGGCAGGTGGT	671
	PCR_891011-R	GAGGTTTTTGATTTTACTTGGAGGT	

Jurkat E6.1 Single-Cell Sorting

Jurkat E6.1 bulk cell populations were subjected to single-cell sorting into 96-well cell culture plates – 2 plates per condition. In advance, 150 µL of complete RPMI medium were added to each well of 96-well cell culture plates. Two millions of cells from each condition were centrifuged at 300 g for 5 min, the supernatants were discarded and cells re-suspended in 500 µL of a Fluorescence Activated Cell Sorting (FACS) buffer (2% FBS, 5 mM EDTA and 0,1% sodium azide in 1x PBS; stored at 4ºC) and then filtered through 0,2 mm cell strainers into FACS tubes. Single-cell sorting was performed using the FACS Aria II Cell Sorter (BD Biosciences). After the single-cell sorting procedure, cells were put at standard culture conditions, with clonal populations being split between two 96-well

plates: one for being used for genotyping (see section above), and one for maintaining in culture in the event of the clone being deletion positive.

Results

Intracellular Distribution of Mcl-1 Varies According to Its Transcript's 3' UTR Length

The discovery made by Mayr and Berkovits[52] showing that the 3' UTR can modulate differential protein localization, prompted us to investigate whether the cellular distribution of the Mcl-1 protein varied when translated from the mRNA containing the short (1418 nt- pA1) or the long (2828 nt - pA2) 3'UTR. Previously, our group has mapped the two *MCL1* APA-derived mRNA isoforms in HeLa cells by 3' RACE (Rapid amplification of cDNA ends). Given that the HeLa cell line is very amenable to work with, such as carrying on overexpression studies through transient transfection, and that E6.1 cells are, in contrast, very difficult to transfect, we decided to perform subcellular localization studies in the HeLa cell line. In order to do so, the following expression constructs featuring a *MCL1* coding sequence (CDS) followed by either the short or the long 3' UTR sequences were made and transfected into HeLa cells: pEGFP *MCL1* CDS pA1 (thereafter referred to as –pA1) and pEGFP *MCL1* CDS pA2 (thereafter referred to as –pA2), with the upstream pA1 mutated. Subcellular localization was analysed by microscopy and the settings of image acquisition in the Sp5 confocal microscope were maintained the same for the two constructs. After being acquired in a 1024x1024 pixel format, the images from the three fluorescence channels (EGFP – EGFP-fused Mcl-1; Mitotracker Red (Mito. Red) – mitochondria and Hoechst –nuclei) were merged using the ImageJ software (Fig. 8, 9).

The Mcl-1 protein expressed from the pEGFP *MCL1* CDS pA1 construct showed a well-defined mitochondrial enrichment, with relatively little cytoplasmic and nuclear localization (Fig. 8a). On the other hand, Mcl-1 expressed from the pEGFP *MCL1* CDS pA2 construct, presented a markedly different localization profile in comparison to the previous construct. Expression of Mcl-1 from this construct is ubiquitous and disperse, showing cytoplasmic and nuclear labelling, and also presenting cytoplasmic speckles (Fig. 9).

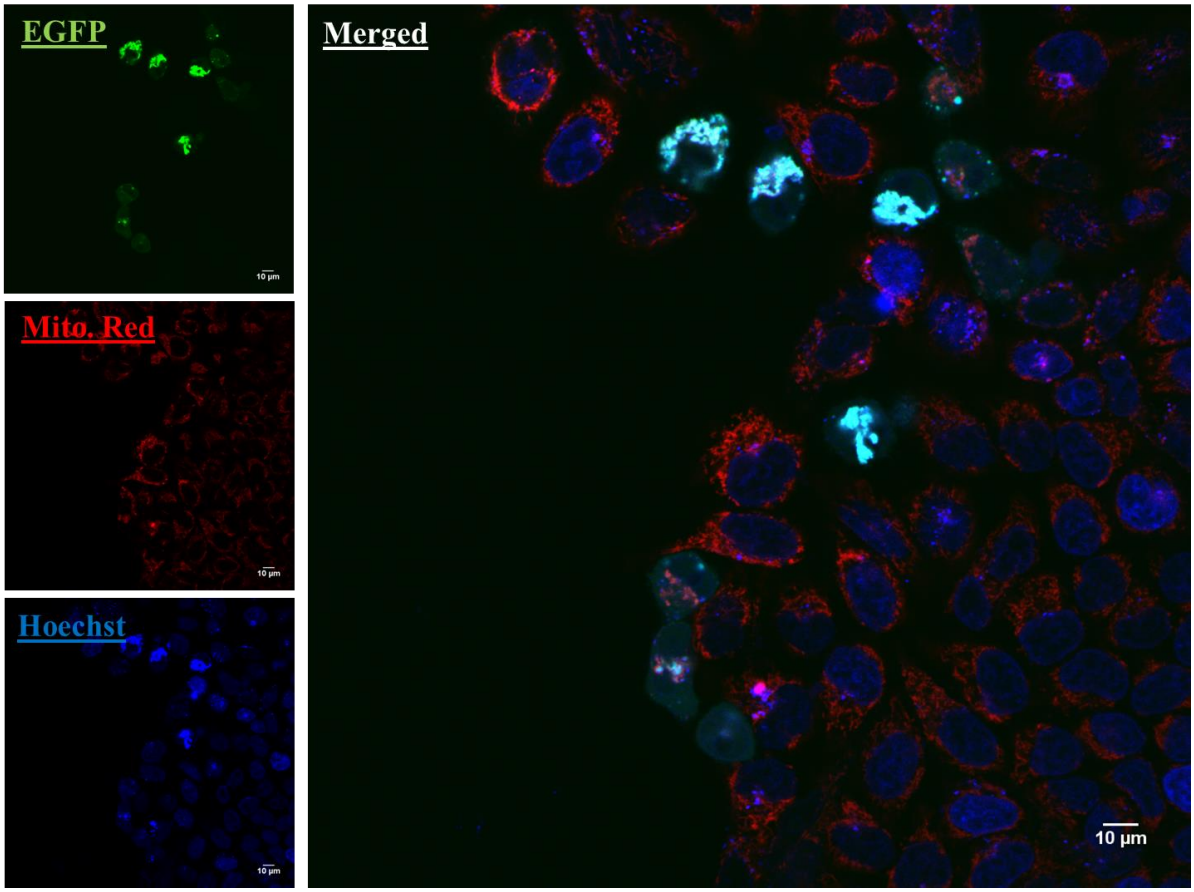

Figure 8 – Mcl-1 cellular localization in HeLa cells, when expressed from the pEGFP MCL1 CDS pA1 construct. On the left column, there are split channels (EGFP: EGFP-fused Mcl-1; Mito. Red: mitochondria; Hoechst: nuclei).

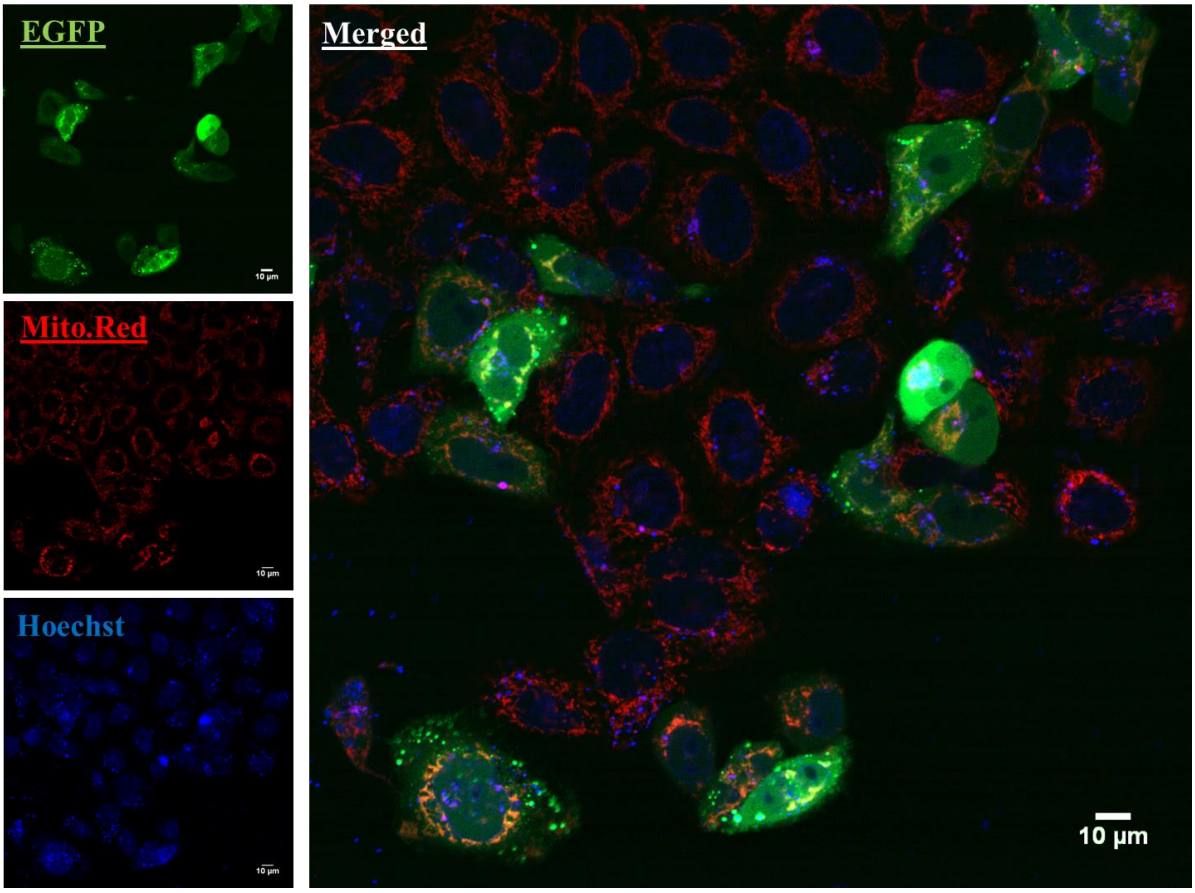

Figure 9 – Mcl-1 cellular localization in HeLa cells, when expressed from the pEGFP *MCL1* CDS pA2 construct. On the left column, there are split channels (EGFP: EGFP-fused Mcl-1; Mito. Red: mitochondria; Hoechst: nuclei.) On the right, there is a merged image of the three channels. In a) the GFP channel fluorescence intensity was maintained according to the parameters set when acquiring images in the Sp5 confocal microscope. The fluorescence intensity in on this image was enhanced using ImageJ to facilitate the visualization of the EGFP-Mcl-1 protein.

Mitochondria Morphology Is Altered Upon Overexpression Of Mcl-1 From The pEGFP MCL1 *CDS pA1 Construct*

In comparison to the non-transfected cells, cells transfected with the pEGFP *MCL1* CDS pA1 construct presented an altered mitochondrial structure, consisting in its shifting from a ring appearance to a more agglomerated phenotype (Fig. 10b). This altered morphology was not observed in cells transfected with the pA2 construct (Fig 9).

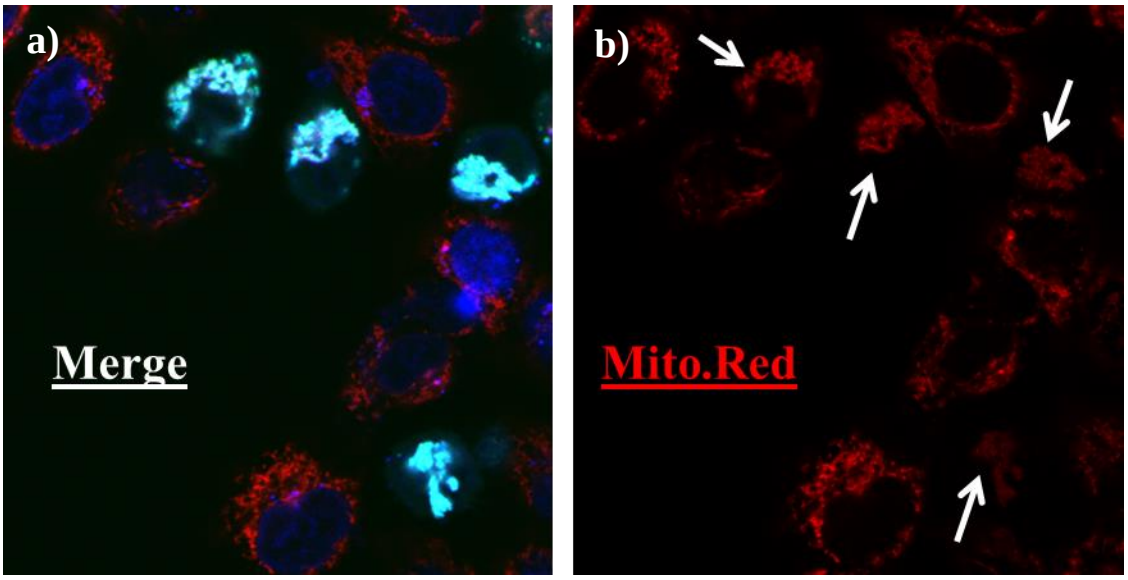

Figure 10 –Mitochondria morphology in HeLa cells changes upon overexpression of Mcl-1 from the pEGFP MCL1 CDS pA1 expression construct. **a)** merged image (EGFP: EGFP-fused Mcl-1; Mito.Red: mitochondria; Hoechst: nuclei). **b)** image depicting mitochondria, using a Mito.Red channel. The transfected cells [white arrows in b)], present a more agglomerated phenotype in comparison to the non-transfected cells, which feature a ring morphology.

The Short pA1 mRNA Isoform Produces More Protein Than pA2

Given that we observed clear differences in EGFP fluorescence intensity between -pA1 and -pA2 EGFP-constructs in the confocal microscope (Fig. 8, 9), we proceeded to EGFP fluorescence quantification (Fig. 11). Firstly, the areas and the mean fluorescence intensity values for each transfected cell were measured using the ROI (Regions Of Interest) manager tool available in the ImageJ software. Thereafter, those mean fluorescence intensity values were corrected by subtracting the mean value of background fluorescence intensity and the obtained values were then multiplied by the area of each cell, with total fluorescence intensity levels being thus quantified for ~50 cells per each condition (-pA1 and -pA2). Afterwards, some outliers were eliminated from each group, with data from a total of 45 cells for each condition being thus used downstream.

In order to quantify the differences in EGFP intensity, the total fluorescence intensity values for 45 cells in the two transfection conditions were uploaded into the GraphPad Prism 7.04 software. Then, an unpaired *t* test with Welch's correction was applied to the datasets, which showed a highly significant difference ($p < 0{,}0001$) between the means of -pA1 (3682572 arbitrary units – a.u.) and -pA2 (1138887 a.u.), Thus, the

mean value of -pA1 total fluorescence intensity is approximately three times higher than the one of -pA2 (Fig. 11).

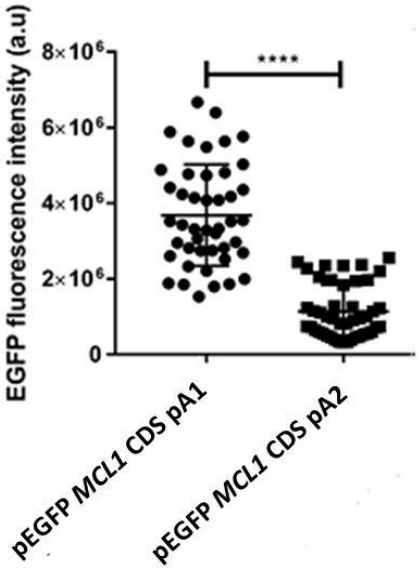

Figure 11 – Fluorescence intensity quantification for the pEGFP MCL1 CDS pA1 and -pA2 constructs. A total of 45 cells in both depicted transfection conditions were quantified, with the fluorescence intensity being measured in arbitrary units (a.u). The difference between the two datasets means was highly significant, being thus depicted by four asterisks (**** $p < 0.001$). For statistical analysis, an unpaired *t* test with Welch's correction was performed. This result is representative of three experiments (n=3).

miR-17 Is Not Responsible for pA2-derived Mcl-1 Expression and Subcellular Localization

Given that we detected a ~3 fold higher fluorescence intensity in the -pA1 construct, we decided to transfect the pEGFP *MCL1* CDS pA2ΔmiR17 expression construct, with a mutated miR-17 binding site, alongside the pA1 mutation. As the miR-17 was previously shown by our group to bind only the long pA2 *MCL1* APA-derived mRNA isoform, we sought to discover whether the fluorescence intensity, which positively correlates with protein levels, would be enhanced upon inhibiting miR-17-mediated regulation of the –pA2 isoform.

However, the difference between the EGFP fluorescence intensity means of -pA2 (1138887 a.u.) and -pA2ΔmiR17 (1308562 a.u.) was not statistically significant (p = 0,2462), when performing an unpaired *t* test with Welch's correction (Fig. 12).

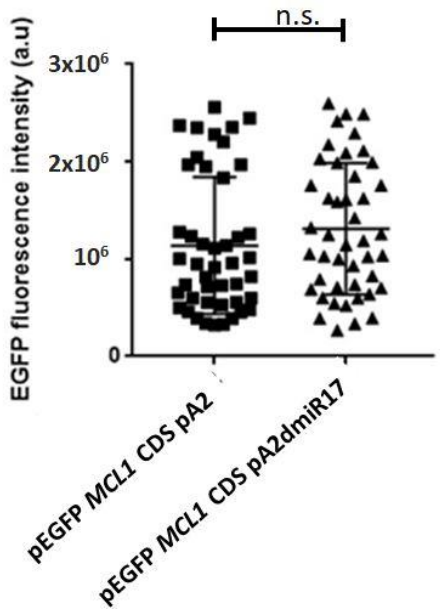

Figure 12 – Fluorescence intensity quantification of the pEGFP MCL1 CDS pA2, and -pA2ΔmiR17 constructs. A total of 45 cells in each of the two depicted transfection conditions were quantified, with the fluorescence intensity being measured in arbitrary units (a.u). An Unpaired t test with Welch's correction statistical analysis was performed, showing that the difference between datasets means was not statistically significant (n.s.).

In addition, Mcl-1 expressed from the - pA2ΔmiR17 construct (Fig. 13) presented a similar subcellular distribution to the one originated from the - pA2 construct (Fig. 9).

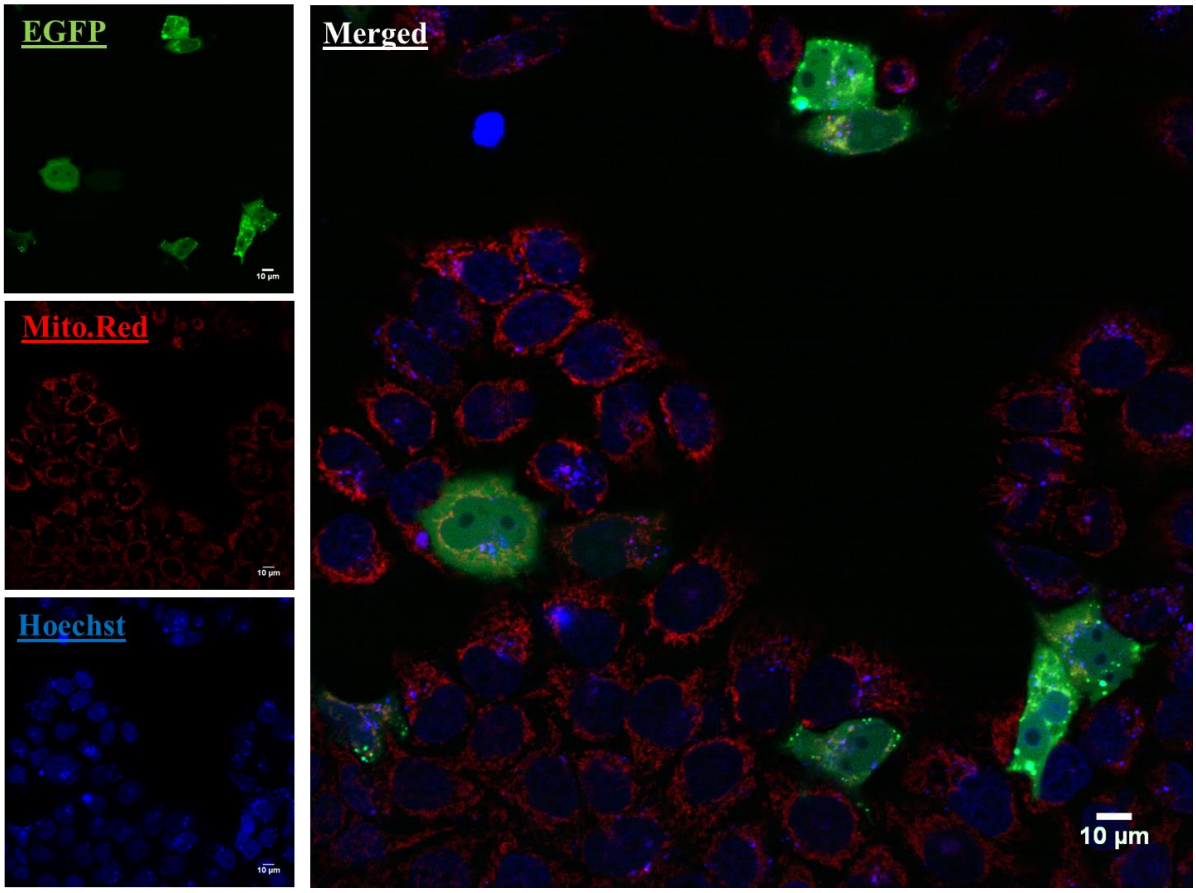

Figure 13 – Mcl-1 subcellular localization in HeLa cells, when expressed from the pEGFP *MCL1* pA2ΔmiR17 construct. On the left column, there are split channels (EGFP: EGFP-fused Mcl-1; Mito. Red: mitochondria; Hoechst: nuclei). On the right, there is a merged image of the three channels. The EGFP channel fluorescence intensity was maintained according to the parameters set when acquiring images on the Sp5 confocal microscope in order to perform correct EGFP quantifications. The fluorescence intensity in this image was enhanced using ImageJ to facilitate the visualization of the EGFP-Mcl-1 protein.

Genotyping of CRISPR/Cas9-subjected Bulk E6.1 Populations and Isolated Clones Proved The Genome Editing to Be Successful

With the aim of studying separately the function of the *MCL1* 3' UTR APA-derived mRNA isoforms, we used an innovative take on the genome engineering CRISPR/Cas9 technology to create stable Jurkat E6.1 cell lines lacking one of the *MCL1* pA signals (pA1 or pA2). In addition, we also applied CRISPR/Cas9 on two other genes involved in *MCL1* APA: INTS9 and CPSF1.

After ten days of puromycin clonal selection of Jurkat E6.1 cells in all eight pLenti –sgRNA conditions (Materials and Methods, Lentiviral Transduction), we obtained stably dividing cells, which integrated the all-in-one lentiviral vector, containing hSpCas9, the chimeric single guide RNA and the puromycin resistance (PuroR) expression cassettes (Materials and Methods, Lentiviral Transduction). Therefore, we proceeded to perform the first genotyping on those bulk cell populations (Fig. 14a and 14b). This was performed by PCR amplification of the genomic regions encompassing the sites of sgRNAs targeting and subsequent analysis of the respective electrophoretic profiles in a 1,5% agarose gel, always on the background of the non-transduced control E6.1 population (Fig. 14a and 14b, c).

The *CPSF1* KO (Fig. 14a, lane s1), *INTS9* KO (Fig. 14a, lane s2), *MCL1*ΔpA1s (single sgRNA approach) (Fig. 14a, lane s3) and *MCL1*ΔpA1d (double sgRNA approach) [Fig. 14a, lane s4], pLenti – sgRNA conditions (Materials and Methods, "Guide RNAs Design Strategy") presented bands migrating a bit slower than the respective E6.1 control bands (Fig. 14a, lanes c). All of the four *MCL1*ΔpA2d pLenti –sgRNA conditions (Fig. 14b, lanes s1 to s4), on the other hand, presented two bands. In each of these four conditions, there was one 671 bp band corresponding to an amplicon size without deletion – the wild-type allele, as in the control (Fig. 14b, lane c), in addition to another lower and much fainter band with a difference of ~200 bp (Fig. 14b, lanes s1 and s3) and ~300 bp (Fig. 14b, lanes s2 and s4), respectively, correspondent to the deletion allele.

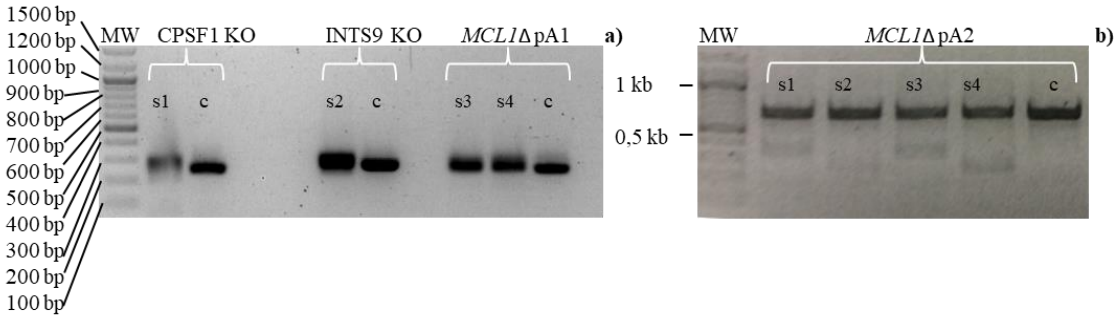

Figure 14 – Jurkat E6.1 CRISPR/Cas9 bulk populations genotyping. Legend: c- control; MW- molecular weight DNA marker; s- sample. **a) s1**: CPSF1 – upstream – up, and downstream – down, guides; **s2**: INTS9 – up and down guides; **s3**: *MCL1*ΔpA1s – single guide; **s4**: *MCL1*ΔpA1d – up and down guides; c – E6.1 **b) s1**: *MCL1*ΔpA2 – up1 and down1 guides; **s2**: *MCL1*ΔpA2 – up1 and down2 guides; **s3**: *MCL1*ΔpA2 – up2 and down1 guides; **s4**: *MCL1*ΔpA2 – up2 and down2 guides; c- E6.1. The guide RNAs list is available on table 1 (forward sequences in capital letters).

The subsequent Sanger sequencing of the 8 samples and the associated controls has confirmed the occurrence of deletions and their specificity, in seven out of eight pLenti – sgRNA conditions, except for *MCL1*ΔpA1s. When the sequences were aligned with the control sequences using the Geneious *software*, in *MCL1*ΔpA2d case ("up1, down1" guide pair – Materials and Methods, Guide RNAs Design Strategy), there was a gap delimited by the respective pair of the sgRNAs (Fig. 15a), whereas for the *CPSF1* KO condition, as expected, there was a frameshift being observed in comparison to the E6.1 control and reference sequences (Fig. 15b), as well as for INTS9 KO (not shown).

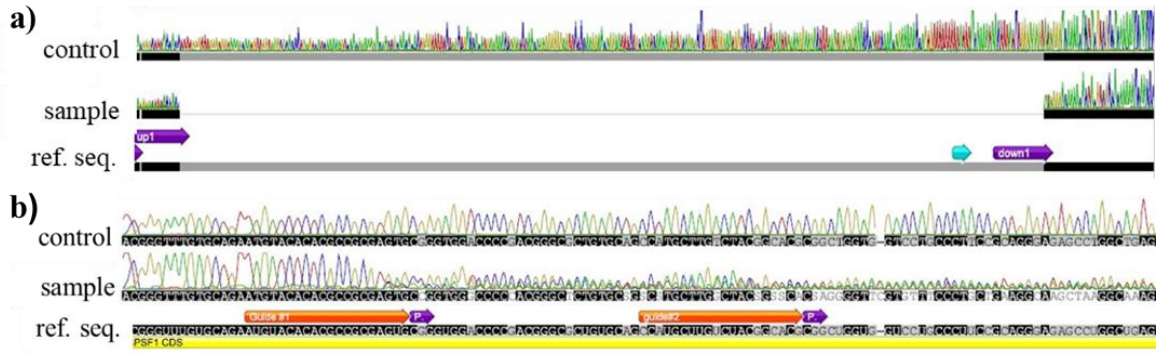

Figure 15 – Sequence alignment of the E6.1 control sequence with the MCL1ΔpA2 **(a)** and CPSF1 KO **(b)** samples, using the Geneious software. **a)** The sequence of the 3'UTR control (non-transduced) E6.1 cells neighbouring region of pA2, top, was aligned with the same region of one of the transduced E6.1 bulk cell populations, middle, and the reference sequence (ref.seq.), bottom, in the program Geneious. A large deletion including pA2 (light blue arrow) is evident. Guides are depicted by purple boxes ("up1, down1" pair). **b)** The sequence of the E6.1 control sequence was aligned with the *CPSF1* KO sample using the Geneious *software*, as in **a)**, with ref.seq. also being depicted, bottom. Instead of a large deletion, there was a frameshift observed in the sequence after guide #1.

Given that all pA2 samples presented the deletion and due to their redundancy, aliquots of bulk populations from the *MCL1*ΔpA2d "up1, down2" and "up2, down1" guide pairs conditions (Materials and Methods, "Guide RNAs Design Strategy) were stored at -80ºC.

Thereafter, single-cell sorting of the 5 out of the remaining 6 conditions was performed (except for *MCL1*ΔpA1s, due to the lack of pA1 deletion in the bulk population genotyping) into 96-well plates – 2 plates per condition. After clonal expansion, genotyping was conducted on isolated clones (Fig. 16), with the intention of obtaining

homozygous clones in each condition. While *CPSF1* KO and the *MCL1*ΔpA1d pLenti –sgRNA conditions have shown a distinct electrophoretic profile from that of the control (Fig. 16a, the first two lanes c), as it has been previously observed for the bulk populations (Fig. 14a), *INTS9* KO clones did not display this difference (Fig. 16a, the third lane c).

As expected, upon subsequent Sanger sequencing, heterozygosity was detected in all *MCL1*ΔpA1d and *CPSF1* clones, except for the *INTS9* KO clones, which presented a wild-type genotype.

Concerning the two *MCL1*ΔpA2d pLenti –sgRNA conditions, we used primers located inside the region of the expected deletion, between the two different sgRNAs pairs (Fig. 16b, lanes s1 and s2). This approach allowed us to identify the wild-type allele, therefore no amplicon was expected after PCR amplification if there was an homozygous deletion.

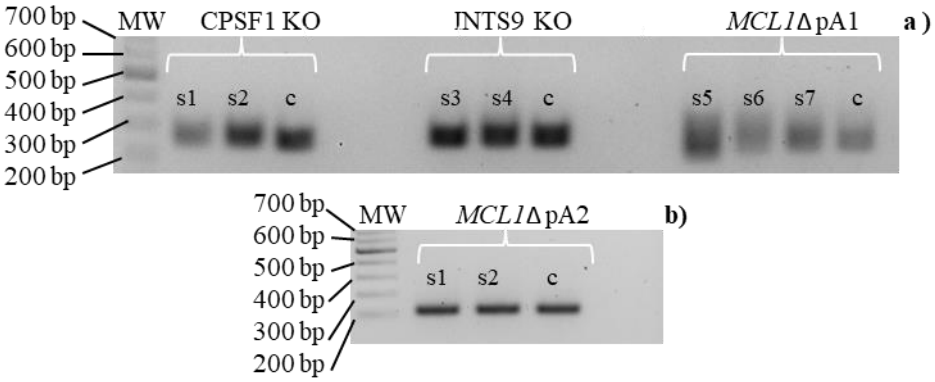

Figure 16 – Single-cell sorting –derived clones genotyping. Legend: c- control; MW- molecular weight DNA marker; s- sample. **a)** s1, s2: CPSF1 KO – up and down guides; s3, s4: INTS9 KO – up and down guides; s5, s6, s7: *MCL1*ΔpA1d – up and down guides; **b)** s1: *MCL1*ΔpA2 up2 and down2 guides; s2: *MCL1*ΔpA2 up1 and down1 guides.

Discussion

MCL-1 Alternative 3'UTRs have a function in protein Subcellular Localization

The subcellular distribution of Mcl-1 was shown to be mainly mitochondrial, in accordance with the its anti-apoptotic function at the OMM[148–150], as well as with its emerging role as a regulator of mitochondrial metabolism and dynamics (fission/fusion) in the mitochondrial matrix[107,154,155]. On the other hand, there were also observations of Mcl-1 nuclear localization[156–159], where Mcl-1 carried on distinct mitochondria-unrelated functions. Some of these functions consisted in the regulation of cell cycle progression when associated to Cdk1[157], or with the mesenchymal-epithelial transition when physically interacting with STAT3[159]. The latter actually leads to an increased rate of apoptosis, as opposed to the conventional anti-apoptotic function of Mcl-1 full-length protein isoform[119].

It has been previously described that the Mcl-1 mitochondrial localization is greatly enhanced by its N-terminus, namely by the first 79 aa featuring a putative mitochondria-targeting sequence[150], as well as relying heavily on the presence of the C-terminal Transmembrane Domain, which allows anchoring to the OMM[150,199]. In accordance, Mcl-1S AS isoform, which lacks the C-terminal Transmembrane Domain is primarily cytosolic[133]. Moreover, the Mcl-1 subcellular localization also relies on its recruitment by protein interaction partners (PIPs), namely the Mcl-1-Noxa complex mitochondrial co-localization[151] and Mcl-1 recruitment to the nucleus by the IEX-1 factor[158]. However, the involvement of non-coding sequences in Mcl-1 localization has not been investigated yet. Our group has previously unveiled the expression of two APA-derived mRNA isoforms in T cells, varying only in the length of their 3' UTRs – the short isoform, which uses the proximal pA signal (pA1) and has a 1418 nt long 3' UTR, and the long isoform, which uses the distal pA (pA2) and has a 2828 nt long 3' UTR[197]. This extensive 3' UTR harbors various potential and identified RBP binding motifs[135–138], apart from miRNAs binding sites[139], allowing for a complex regulation of *MCL1* transcripts fate, such as mRNA stability and translation.

In addition to these two aspects of modulation of gene expression, the role of the 3' UTR in protein function in subcellular localization has been previously demonstrated by Mayr and Berkovits for CD47, who presented the 3' UTR-dependent protein localization (UDPL) model, using a range of human cell lines[52]. Their work uncovered that HuR only binds to the long *CD47* 3' UTR mRNA isoform, where it serves as an RBP adaptor for

SET, the latter being transferred during translation from the 3' UTR onto the CD47 nascent cytoplasmic domains, thereupon recruiting RAC1 (Fig. 2). Thereafter, active RAC1 translocates the whole protein complex to the plasma membrane (PM), where high levels of CD47 are needed to fully protect cells from phagocytosis[52].

On the other hand, the short *CD47* 3' UTR APA-derived mRNA isoform, which cannot bind by HuR, is translated into a protein that while not being targeted to the PM, accounts for CD47 cytoplasmic localization. However, unlike PM CD47, cytoplasmic CD47 is pro-apoptotic, restoring apoptosis in gamma-irradiated CD47-defficient Jurkat cells (JinB8), upon exogenous expression from the short *CD47* 3' UTR APA-derived mRNA isoform, but not from the long[52]. Thus, thanks to the regulatory landscape provided by the 3' UTR, both a differential localization and an independent expression of two proteins is being achieved, which while bearing the same aminoacid sequence, perform two distinct functions.

Inspired by the elegant model proposed by Mayr and Berkovits[52], we sought to discover whether the two *MCL1* transcript isoforms with alternative 3' UTRs, may mediate a similar sort of regulation, namely protein partners recruitment, and thus, target the Mcl-1 protein to different subcellular compartments (mitochondria, cytosol, nucleus). In order to do so, we performed overexpression and localization assays in HeLa cells using the pEGFP *MCL1* CDS pA1 and the pEGFP *MCL1* CDS pA2 constructs. Indeed, like Mayr and Berkovits[52], we observed a differential Mcl-1 subcellular localization dependent on the 3'UTR: predominantly mitochondrial for the *MCL1* CDS pA1, although with very low cytoplasmic and nuclear expression (Fig. 8), and ubiquitous for *MCL1* CDS pA2 (Fig. 9), encouraging us to further explore the possibility of Mcl-1 regulation according to the UDPL model. Given that the *MCL1* CDS pA1 mitochondrial localization was consistent with the previous reports on Mcl-1 localization using expression constructs containing *only* the coding sequence[149–151], we deemed that the pA1 3' UTR was not essential for Mcl-1 mitochondrial targeting. However, our first hint was that the pA2 3' UTR could specify for alternative Mcl-1 localizations, other than mitochondrial, namely: nucleus and cytoplasm, thereby dissolving the mitochondria-shifted protein accumulation. According to the bioinformatics analysis using the RBPmap Version 1.1 (http://rbpmap.technion.ac.il/), there are five putative RBP binding motifs in the pA2 3' UTR sequence that do not overlap with pA1 3' UTR. These are binding motifs for FXR1 (fragile X mental retardation

syndrome-related protein 1), FXR2, PCBP2 (Poly(rC)-binding protein 2), RBM3 (RNA-binding protein 3) and SFPQ (splicing factor, proline- and glutamine-rich), whose RBPs expression in T cells was confirmed using the EMBL-EBI Expression Atlas. However, when assessing the five RBPs' gene ontology there was no hit for a role in protein localization, but rather for regulation of protein translation rate and stability, such as for FXR1[200] and RBM3[201]. Still, these RBPs may themselves be responsible for modulating *MCL1* CDS pA2-derived Mcl-1 localization, or may recruit protein effectors, acting as RBP adaptors[53]. However, given that the 3' UTR region harbors numerous repeats and stretches of identical nucleotides (A/U), it may also form secondary structures that may affect the accessibility of certain miRNAs[202] or RBPs binding sites[53,54,203]. Thus, when looking into candidate proteins involved in Mcl-1 UDPL regulation, in fact, one should consider the whole sequence and not only the non-overlapping part.

Still in compliance with the Mayr and Berkovits' UDPL model hypothesis[52], the differential protein localization could stem from a difference in protein import into the mitochondria. Slower mitochondrial import has already been reported for exogenously expressed Mcl-1 with a truncated N-terminus[150,199]. Even though the two Mcl-1 proteins (*MCL1* CDS pA1 and pA2- derived) bear an identical amino acid sequence, only differing in their 3' UTRs, they still may engage with different protein partners recruited by their alternative 3' UTRs[52,53]. Thus, one type of PPI could target the Mcl-1 protein complex more efficiently to the mitochondria, and/or allow it to be more readily imported. Furthermore, the previously reported mitochondria co-localization event promoted by the establishment of the Noxa/ PMA1P - Mcl-1 PPI[151] could hypothetically be the one modulated by UDPL. Even though the protein expressed from the *MCL1* CDS still engages with Noxa[151] and is thus targeted and imported into mitochondria, the UDPL model can potentially speed up this interaction.

Another possible protein interactors involved in mediating the import process are heat shock proteins (HSPs), some of which were recently unveiled as non-conventional RBPs, capable of binding RNA through their intrinsically disordered regions (IDRs)[54]. There are two chaperone systems involved in mitochondrial transport: Hsp90 (Heat shock protein 90)/p23 and Hsc70 (Heat shock cognate 70 kDa)/Hsp40[204]. In both systems, heat shock proteins bind newly translated mitochondria-targeted protein precursors, in the

cytosol, and escort them to the mitochondria in a translocation-competent, unfolded, state[204]. There, they facilitate mitochondrial entry upon promoting mitochondria-targeted protein interaction with the TOM (translocase of the outer membrane) receptors[204], with Mcl-1 having been shown to interact with the mitochondrial import receptor Tom70, through an N-terminus internal motif[199]. According to Biogrid 3.4 site (https://thebiogrid.org), Mcl-1 directly interacts with three HSPs: HSP1A1, HSPA4 and HSPA8, with HSP1A1, in particular, being recently identified as an unconventional RBP[54]. Thus, they could interact with Mcl-1 through their recruitment to the 3' UTR of *MCL1* transcripts, which could be preferential of the short pA1 isoform. Alternatively, HSPs could be recruited by other RBP adaptors to the nascent protein, modulating its downstream fate.

In accordance with the previous finding of our group showing by reporter assays that the pA1 isoform produces more protein, the pEGFP *MCL1* CDS pA1 construct gave rise to a significantly higher level of EGFP fluorescence intensity (Fig. 11), with pA1 producing ~3-fold more protein than *MCL1* CDS pA2. This more efficient protein production from the short isoform was most likely related to the evasion of cytoplasmic regulation by miRNAs, different kinds of RBPs and through destabilizing *cis*-elements present and/or only made available in the long 3' UTR. Thus, we sought to find whether miR-17 was responsible for this downregulation of Mcl-1 expression in *MCL1* CDS pA2. However, the *MCL1* CDS pA2ΔmiR17 construct, with its mutated miR17 binding site[197], located inside the non-overlapping region of the pA2 3' UTR, did not render a significant enhancement of Mcl-1 expression when compared to the pEGFP *MCL1* CDS pA2 construct (Fig. 12). That finding only proved the complexity of cell regulatory processes, with multiple cytosolic events accounting for negative modulation of mRNA expression.

Moreover, overexpression of *MCL1* CDS pA1 seemed to affect mitochondria morphology, namely shifting it from a ring appearance to a more agglomerated phenotype (Fig. 10b). This was opposed to *MCL1* CDS pA2, but in compliance with the previous studies linking Mcl-1 with modulation of mitochondria morphology. Thus, knockdown of Mcl-1 was recently reported to promote mitochondria organelles fusion and elongation in human Embryonic Stem Cells (hESCs)[155]. Mcl-1 high expression levels, on the other hand,

were linked with a fragmented mitochondrial network (fission) in induced Pluripotent Stem Cells (iPSCs), in comparison to the parental fibroblasts[155]. Taking those findings into account, in the future, it would be interesting to investigate a possible role for *MCL1* 3' UTR in shaping mitochondria morphology.

Overall, the EGFP reporter localization studies revealed that we are currently only in the beginning of unraveling the complex regulation to which *MCL1* is subjected. However, we are now confident that MCL1 3' UTR regulatory potential extends beyond the regulation of its transcript translation efficiency, affecting also the protein fate, including subcellular localization.

CRISPR/Cas9 Genome Editing Of Jurkat E6.1 Cells

Although the GFP reporter gene assay provides one of the most expeditious ways to tag and explore the localization profile of one's protein of interest (POI), it has its intrinsic caveats. For instance, the expression of exogenous protein at non-physiological levels may disturb cellular functions, altering the cell phenotype[205,206]. In addition, overexpression can alter trafficking or post-translational modification of the protein N-terminally fused with a 27 kDa GFP, with a tag being capable of interfering with protein folding and establishment of protein–protein interactions[205]. Thus, data gathered from GFP reporter gene assays may differ from what can be observed in native conditions.

Thus, we sought to carry on our research in a clean, unbiased and physiological framework, by applying the CRISPR/Cas9 genome editing tool on Jurkat E6.1 T cells, with the aim of dissecting separately the *in vivo* functions of the *MCL1* pA1 (~1,5 kb) and pA2 (~2,8 kb) alternative 3' UTRs, in their physiological context. In addition, we asked whether the *knock-out* of two relevant genes involved in RNA processing, INTS9 and CPSF1, would impact *MCL1* polyadenylation and/or alternative polyadenylation outcome, as well as cell viability. .

The gene delivery systems can be either viral- (cell transduction) or non-viral-mediated (cell transfection). Although advances and optimization efforts in cell transfection have brought promising results in recent years, such as in electroporation[207–209], DNA transfection may limit the use of multiplex genome engineering due to the associated toxicity, as described for primary T cells upon DNA electroporation[210]. Viral-mediated delivery does still present the higher efficiency[211], despite the fact that the

integration of a viral gene into a host cell genome poses serious safety concerns[212,213] as well as experimental drawbacks (e.g. the risk of insertional mutagenesis[213]). Given that Jurkat E6.1 cells, have been proven as hard-to-transfect[208,214], and the advantages of viral-mediated delivery, we chose lentiviral transduction, using the lentiCRISPR v2 vector, to deliver our all-in-one pLenti –sgRNA constructs (gRNAs listed in table 1) into E6.1 cells.

The occurrence of E6.1 genome editing in seven out of eight pLenti –sgRNA conditions in bulk cell populations, corroborated by PCR (Fig. 14a and 14b) and Sanger sequencing (Fig. 15a and 15b) proved the efficiency of our lentiviral transduction. The observed mixed populations of wild type alleles and the ones harboring indels or deletions (Fig. 14 and 15) suggested three possible outcomes of our sgRNAs genome targeting approaches, namely: the existence of wild type cells and the ones with the two of their respective alleles being mutated, the existence of heterozygous cells and, much more probably, the overlapping of the two scenarios. The latter hypothesis was confirmed upon obtaining isolated clonal populations for each condition, and genotyping, which rendered heterozygous and wild-type genotypes (Fig. 16a and 16b).

Overall, we obtained only a few clones in each pLenti – sgRNA condition (Fig. 16) upon single-cell sorting into 96-well cell-culture plates, which was not surprising. This was because, on one hand, the scale of our single-cell sorting experiment (two plates per condition) was small and, on the other hand, because the Jurkat cells, including E6.1, respond negatively to low density/isolation, appearing to need growth factors produced by neighbor cells.

Nevertheless, the majority of our clones presented a deletion, but all of those were heterozygous. This could be because the Cas9 endonuclease mediated genome editing is a stochastic event, whose outcome is therefore unpredictable[215], in spite the good efficiency (score) of the designed sgRNAs (ranging from 70 in *MCL1*ΔpA2d to 96 in CPSF1 KO) [table. 1]. Even though heterozygosity was achieved for our clones, it was our aim to attain homozygosity to allow functional assays to be performed.

A large number of APA factors are required for appropriate polyadenylation processing to occur, including cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CSTF) and cleavage factor I and II (CFI and CFII)

complexes, the poly(A) polymerase, as well as a panoply of other associated factors[43,216].

Some of these basal polyadenylation factors have already been described to have roles in APA as well. CSTF2/CstF64, which is an essential polyadenylation factor binding the GU-rich DSEs[10], was demonstrated to act as an important regulator of 3′ UTR shortening in B cell differentiation[60] and across several cancer types [e.g. bladder urothelial carcinoma (BLCA), head and neck squamous cell carcinoma (LUSC)][217]. CstF-64 and CFI(m) proteins have much higher positional specificity around the PAS most frequently used than CPSF [218]. However, the latter plays a major role in recruiting CstF-64 to RNAs[219].

The CPSF1/CPSF160 factor, while not being shown to be involved in APA so far, has already been described as an important regulator of alternative splicing, either inhibiting alternative exon inclusion, such as in the Interleukin 7 receptor – *IL7R*[19], in T cells[19], or promoting it, as shown for the Androgen Receptor – *AR*[21], in Castration Resistant Prostate Cancer (CRPC) cells[21] and also at the genome wide level [220]. Being mindful of the recent study made by Sun and colleagues[44], which uncovered that CPSF1 functions as a recruitment platform for CPSF-30 and WDR33 and primes them for high-affinity binding to AAUAAA, allows revisiting the earlier established roles for CPSF1. The latter makes it clear that while CPSF1 does not interact with pre-mRNA directly, it is still crucial for the assembly of the CPSF complex on internal polyadenylation signals, and therefore its interplaying with the spliceosome machinery.

With CPSF1 being a basal polyadenylation factor, and armed with knowledge available to date concerning CPSF1 regulatory function, we sought to pursue its possible implication in *MCL1* APA outcome. In order to achieve this, we planned a two-allele KO for its gene in E6.1 cells, through lentiviral transduction of all-in-one pLenti – sgRNA constructs.

The fact that in the beginning of the puromycin selection, there were high mortality levels across all of the eight CRISPR/Cas9 conditions (Materials and Methods, Lentiviral Transduction) was not surprising given the incomplete efficiency of the transduction approach, with only cells which incorporated the lentiCRISPR v2 constructs expressing the PuroR gene and hence surviving. However, the mortality was notoriously higher in the beginning for the *CPSF1* KO condition, having stabilized only after a week of puromycin

selection. This observation suggests that the homozygous CPSF1 *knockout* was lethal for the cells, highlighting even on the background of the overall cell mortality levels due to the Puro-induced cell death of the non-transduced part of the bulk population, in comparison to other conditions. The following stabilization was probably due to the survival and proliferation of heterozygous clones. In fact, we have detected those upon bulk population and single-cell sorting isolated clones genotyping (Fig. 14a and 16a, respectively).

Integrator (INT) is a transcriptional regulatory complex associated with the CTD of the RNAPII[221], being endowed with a core catalytic RNA endonuclease activity which is required for the 3' end processing of both uridine-rich small nuclear RNAs (UsnRNAs)[221] and of transcripts derived from distal regulatory elements (enhancers) – enhancer RNAs (eRNAs)[222], the latter being involved in spatiotemporal regulation of gene expression in metazoan species[223–226]. Very recently, it was discovered that the Integrator most critical module, containing the endonuclease activity, was composed of the Integrator subunit 4 (INTS4) interacting with the INTS9/11 (endonuclease)[227].

By cleaving its target transcripts, Integrator thus promotes their processing to a mature form, which is intimately coupled with transcriptional termination[221]. In addition, it was revealed that the Integrator complex plays a termination function at diverse classes of target genes, including replication-dependent histone mRNAs and genes with polyadenylated transcripts[198]. Concerning the latter, Integrator binds the promoter proximal sites, which positively regulates gene expression. Furthermore, Integrator plays a critical role in both initiation and the release of paused RNAPII at immediate early genes (IEGs) following transcriptional activation by epidermal growth factor (EGF) in human cells, recruiting the super elongation complex (SEC) and thus allowing productive transcription elongation[228]. The surprising spectrum of Integrator functions in gene expression has been reviewed in 2015[229].

Given that our group preliminary findings have suggested that INTS9 was implicated in *MCL1* APA in T cells, with its knockdown affecting the expression of the *MCL1* pA2 isoform,we sought to pursue this new possible role for INTS9, having planned a two-allele KO for its gene in E6.1 cells.

In the beginning of puromycin selection, the mortality levels for *INTS9* KO condition were not as striking as those for *CPSF1*, being similar to the *MCL1*ΔpA1s/ pA1d

and *MCL1*ΔpA2d conditions. This observation was suggestive of a non-essential role of INTS9 in cell survival, given that there is no INTS9 function redundancy with other proteins (e.g. other integrator subunits). Although we have found heterozygous clones in the bulk population upon genotyping, no homozygous, nor heterozygous clones have so far been obtained for the INTS9 KO condition due to the low numbers of clones having survived the single-cell sorting.

Although it is difficult to apply CRISPR/Cas9 in the 3' UTR because of its high A/T content due to various AU-rich elements[230], while the PAM sequence trinucleotide needs to have at least two "G", this technology has been previously successfully applied to target 3' UTRs. For instance, Song and colleagues performed a deletion of the tyrosinase (Tyr) 3' UTR in rabbit zygotes[231] and Zhao and colleagues deleted sequences corresponding to 3' UTRs of nine chemokines in Human bronchial epithelial BEAS-2B.tTA cells[232], with a dual sgRNA approach being applied in both cases.

However, to our knowledge, genome editing has not been applied so far to exclusively delete pA signals in the 3' UTR. Therefore, we decided to adopt an innovative take on the CRISPR/Cas9 technology. Our attempts to delete a large genomic region encompassing the pA2 signal (200-300 bp), as well as a smaller region including the pA1 signal (~70 nt), were similar to the previous approaches of Song and colleagues[231] and Zhao and colleagues[232], relying on double sgRNAs targeting. However, the single sgRNA targeting of a sequence near the pA1, (with a predicted cleavage site being only 5 nt apart from the hexamer), relied on the NHEJ pathway to introduce mutations in the pA1 signal sequence, being thus an unprecedented approach. Our intention of deleting a small sequence including the pA1 signal was dictated by the fact that we could not delete an area that was too large, as it may contain regulatory sequences needed for the pA2 mRNA isoform.

We screened for the presence of a wild-type allele in *MCL1*ΔpA2d because we were primarily interested in obtaining homozygous clones and because we had primers available matching the region delimited by the sgRNA pairs. This approach has proven efficient, allowing us to rule out the two *MCL1*ΔpA2d clones straight away as immediate candidates for the functional assays.

Thus, our CRISPR/ Cas9 approaches are yet to realize their full potential, which will only be possible after obtaining the homozygous isolated clones, on which we will perform the functional assays.

In conclusion, the successful establishment in our lab of the complex protocol involved in CRISPR/Cas9 mediated genome editing, ranging from the *in silico* gRNAs design to lentiviral constructs assembly and lentiviral transduction of difficult-to-transfect cells is a big achievement.

Future Perspectives

Mechanisms Of Mcl-1 Subcellular Localization

The next step in unraveling the regulatory landscape of the two *MCL1* 3 UTR APA-derived mRNA isoforms would be to look into the RBPs that bind their alternative 3' ends. Firstly, it would be interesting to knock down some of the RBPSs identified in the RBPmap binding the pA2-isoform specific 3' UTR part of *MCL1* (see section above), and see whether it would affect Mcl-1 localization and/or translation efficiency.

In parallel, given that the mass spectrometry (MS) technology is not feasible to apply to study proteins bound to large RNA sequences, we would need to create various alternative large deletions in the 3' UTR and then perform mass spectrometry to uncover the bound RBPs. Afterwards, it would be possible to perform a UV cross-linking and immunoprecipitation of the *MCL1* 3' UTR-derived ribonucleoparticle (CLIP) using antibodies against the RBPs identified by MS, thus confirming their binding status.

CRISPR/Cas9: The In Vivo Function Of MCL-1 Alternative 3'UTRs And Identification Of MCL1 APA Regulators

In our future work, we will proceed to scale up the single-cell sorting experiment, using approximately ten 96-well plates per each CRISPR/Cas9 condition. Then, we will genotype the isolated clones, screening for homozygous ones. In parallel, we will perform a second round of lentiviral transduction with the pLenti – sgRNA constructs-containing lentiviruses on the already obtained heterozygous clones, in case the homozygosity will continue failing to be achieved upon scaling up the experiment.

Finally, we will perform functional assays on homozygous clones for *INTS9* and *CPSF1 knockouts* and for *MCL1*ΔpA1 and *MCL1*ΔpA2, such as localization of endogenous Mcl-1, which would be a more physiological and thus biologically significant approach in comparison to reporter/overexpression assays conducted with cell transfection. In addition, we will assess the levels of apoptosis in our engineered E6.1 populations (annexinV/propidium iodide assays).

These experiments will further promote our understanding of the *MCL1* 3' UTR-APA-derived mRNA isoforms regulation, by providing us with opportunities to delve deeper into its subjacent mechanisms and molecular players.

Bibliography:

1. Gene expression - Latest research and news | Nature. https://www.nature.com/subjects/gene-expression. Accessed September 17, 2017.
2. Crick F. Central dogma of molecular biology. *Nature*. 1970. http://link.springer.com/article/10.1038/227561a0. Accessed September 17, 2017.
3. Kiełbasa SM, Vingron M, Smit A, Zhang Z, Baertsch R. Transcriptional Autoregulatory Loops Are Highly Conserved in Vertebrate Evolution. Isalan M, ed. *PLoS One*. 2008;3(9):e3210. doi:10.1371/journal.pone.0003210.
4. Wu C -t., Morris JR. Genes, Genetics, and Epigenetics: A Correspondence. *Science (80-)*. 2001;293(5532):1103-1105. doi:10.1126/science.293.5532.1103.
5. Ares M, Proudfoot NJ. The Spanish connection: Transcription and mRNA processing get even closer. *Cell*. 2005;120(2):163-166. doi:10.1016/j.cell.2005.01.002.
6. Moore MJ, Proudfoot NJ. Review Pre-mRNA Processing Reaches Back to Transcription and Ahead to Translation. *Cell*. 2009;136(4):688-700. doi:10.1016/j.cell.2009.02.001.
7. Bentley DL. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet*. 2014;15(3):163-175. doi:10.1038/nrg3662.
8. Keegan LP, Gallo A, O'Connell MA. The many roles of an RNA editor. *Nat Rev Genet*. 2001;2(11):869-878. doi:10.1038/35098584.
9. Lutz CS, Moreira A. Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. *Wiley Interdiscip Rev RNA*. 2011;2(1):22-31. doi:10.1002/wrna.47.
10. Di Giammartino DC, Nishida K, Manley JL. Mechanisms and Consequences of Alternative Polyadenylation. *Mol Cell*. 2011;43(6):853-866. doi:10.1016/j.molcel.2011.08.017.
11. Proudfoot NJ. Ending the message : poly (A) signals then and now. *Genes Dev*. 2011;25:1770-1782. doi:10.1101/gad.17268411.Freely.
12. Shi Y. Alternative polyadenylation: new insights from global analyses. *RNA*. 2012;18(12):2105-2117. doi:10.1261/rna.035899.112.
13. Danckwardt S, Hentze MW, Kulozik AE. 3′ end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J*. 2008;27(3):482-498. doi:10.1038/sj.emboj.7601932.
14. Calvo O, Manley JL. Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev*. 2003;17(11):1321-1327. doi:10.1101/gad.1093603.
15. Manley JL. Nuclear coupling: RNA processing reaches back to transcription. *Nat Struct Biol*. 2002;9(11):790-791. doi:10.1038/nsb1102-790.
16. Proudfoot N. New perspectives on connecting messenger RNA 3′ end formation to transcription. *Curr Opin Cell Biol*. 2004;16(3):272-278. doi:10.1016/j.ceb.2004.03.007.
17. Buratowski S. Connections between mRNA 3′ end processing and transcription termination. *Curr Opin Cell Biol*. 2005;17(3):257-261. doi:10.1016/j.ceb.2005.04.003.
18. Hilgers V. Alternative polyadenylation coupled to transcription initiation: Insights from ELAV-

mediated 3’ UTR extension. *RNA Biol*. 2015;12(9):918-921. doi:10.1080/15476286.2015.1060393.

19. Evsyukova I, Bradrick SS, Gregory SG, Garcia-Blanco MA. Cleavage and polyadenylation specificity factor 1 (CPSF1) regulates alternative splicing of interleukin 7 receptor (IL7R) exon 6. *RNA*. 2013;19(1):103-115. doi:10.1261/rna.035410.112.
20. Misra A, Green MR. From polyadenylation to splicing: Dual role for mRNA 3’ end formation factors. *RNA Biol*. 2016;13(3):259-264. doi:10.1080/15476286.2015.1112490.
21. Van Etten JL, Nyquist M, Li Y, et al. Targeting a Single Alternative Polyadenylation Site Coordinately Blocks Expression of Androgen Receptor mRNA Splice Variants in Prostate Cancer. *Cancer Res*. 2017;77(19):5228-5235. doi:10.1158/0008-5472.CAN-17-0320.
22. Alén C, Kent NA, Jones HS, O’Sullivan J, Aranda A, Proudfoot NJ. A Role for Chromatin Remodeling in Transcriptional Termination by RNA Polymerase II. *Mol Cell*. 2002;10(6):1441-1452. doi:10.1016/S1097-2765(02)00778-5.
23. Lee K, Blobel GA. Chromatin architecture underpinning transcription elongation. *Nucleus*. 2016;7. doi:10.1080/19491034.2016.1200770.
24. Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet*. 2008;9(11):843-854. doi:10.1038/nrg2438.
25. Sachs A. The role of poly(A) in the translation and stability of mRNA. *Curr Opin Cell Biol*. 1990;2(6):1092-1098. doi:10.1016/0955-0674(90)90161-7.
26. Guhaniyogi J, Brewer G. Regulation of mRNA stability in mammalian cells. *Gene*. 2001;265(1-2):11-23. doi:10.1016/S0378-1119(01)00350-X.
27. D’Ambrogio A, Nagaoka K, Richter JD. Translational control of cell growth and malignancy by the CPEBs. *Nat Rev Cancer*. 2013;13(4):283-290. doi:10.1038/nrc3485.
28. Natalizio BJ, Wente SR. Postage for the messenger: designating routes for nuclear mRNA export. *Trends Cell Biol*. 2013;23(8):365-373. doi:10.1016/j.tcb.2013.03.006.
29. Kobayashi M, Sorensen PW, Stacey NE. Hormonal and pheromonal control of spawning behavior in the goldfish. *Fish Physiol Biochem*. 2002;26(1):71-84. doi:10.1023/A:1023375931734.
30. Jalkanen AL, Coleman SJ, Wilusz J. Determinants and implications of mRNA poly(A) tail size--does this protein make my tail look big? *Semin Cell Dev Biol*. 2014;34:24-32. doi:10.1016/j.semcdb.2014.05.018.
31. Beaudoing E. Patterns of Variant Polyadenylation Signal Usage in Human Genes. *Genome Res*. 2000;10(7):1001-1010. doi:10.1101/gr.10.7.1001.
32. Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D. Patterns of variant polyadenylation signal usage in human genes. *Genome Res*. 2000;10(7):1001-1010. http://www.ncbi.nlm.nih.gov/pubmed/10899149. Accessed May 30, 2018.
33. Tian B, Hu J, Zhang H, Lutz CS. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res*. 2005;33(1):201-212. doi:10.1093/nar/gki158.
34. Tian B, Graber JH. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip Rev RNA*. 2012;3(3):385-396. doi:10.1002/wrna.116.
35. Zhang H, Rigo F, Martinson Correspondence HG. Poly(A) Signal-Dependent Transcription

Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. 2015. doi:10.1016/j.molcel.2015.06.008.

36. Darmon SK, Lutz CS. Novel upstream and downstream sequence elements contribute to polyadenylation efficiency. *RNA Biol*. 2012;9(October):1255-1265. doi:10.4161/rna.21957.
37. Nunes NM, Li W, Tian B. A functional human Poly (A) site requires only a potent DSE and an A-rich upstream sequence. 2010;29(9):1523-1536. doi:10.1038/emboj.2010.42.
38. Zarudnaya MI, Kolomiets IM, Potyahaylo AL, Hovorun DM. Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res*. 2003;31(5):1375-1386. http://www.ncbi.nlm.nih.gov/pubmed/12595544. Accessed May 30, 2018.
39. Shi Y, Manley JL. The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev*. 2015;29(9):889-897. doi:10.1101/gad.261974.115.
40. Moreira A, Takagaki Y, Brackenridge S, Wollerton M, Manley JL, Proudfoot NJ. The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. *Genes Dev*. 1998;12(16):2522-2534. http://www.ncbi.nlm.nih.gov/pubmed/9716405. Accessed September 21, 2017.
41. Hall-Pogar T, Zhang H, Tian B, Lutz CS. Alternative polyadenylation of cyclooxygenase-2. *Nucleic Acids Res*. 2005;33(8):2565-2579. doi:10.1093/nar/gki544.
42. Tian B, Manley JL. Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem Sci*. 2013;38(6):312-320. doi:10.1016/j.tibs.2013.03.005.
43. Neve J, Patel R, Wang Z, Louey A, Furger AM. Cleavage and polyadenylation: Ending the message expands gene regulation. *RNA Biol*. 2017;14(7):865-890. doi:10.1080/15476286.2017.1306171.
44. Sun Y, Zhang Y, Hamilton K, et al. Molecular basis for the recognition of the human AAUAAA polyadenylation signal. *Proc Natl Acad Sci*. 2018;115(7):E1419-E1428. doi:10.1073/pnas.1718723115.
45. Lemay J-F, Lemieux C, St-André O, Bachand F. Crossing the borders: poly(A)-binding proteins working on both sides of the fence. *RNA Biol*. 7(3):291-295. http://www.ncbi.nlm.nih.gov/pubmed/20400847. Accessed September 16, 2017.
46. Kuehner JN, Pearson EL, Moore C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol*. 2011;12(5):283-294. doi:10.1038/nrm3098.
47. Proudfoot NJ. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science*. 2016;352(6291):aad9926. doi:10.1126/science.aad9926.
48. Vorlová S, Rocco G, LeFave CV, et al. Induction of Antagonistic Soluble Decoy Receptor Tyrosine Kinases by Intronic PolyA Activation. *Mol Cell*. 2011;43(6):927-939. doi:10.1016/J.MOLCEL.2011.08.009.
49. Hughes TA. Regulation of gene expression by alternative untranslated regions. *Trends Genet*. 2006;22(3):119-122. doi:10.1016/j.tig.2006.01.001.
50. Al-Ahmadi W, Al-Ghamdi M, Al-Haj L, Al-Saif M, Khabar KSA. Alternative polyadenylation variants of the RNA binding protein, HuR: abundance, role of AU-rich elements and auto-

Regulation. *Nucleic Acids Res*. 2009;37(11):3612-3624. doi:10.1093/nar/gkp223.

51. Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett*. 2008;582(14):1977-1986. doi:10.1016/j.febslet.2008.03.004.
52. Berkovits BD, Mayr C. Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. *Nature*. 2015;522(7556):363-367. doi:10.1038/nature14321.
53. Mayr C. Regulation by 3 -Untranslated Regions. 2017;(August):171-194. doi:10.1146/annurev-genet-120116-024704.
54. Hentze MW, Castello A, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol*. 2018. doi:10.1038/nrm.2017.130.
55. Derti A, Garrett-Engele P, Macisaac KD, et al. A quantitative atlas of polyadenylation in five mammals. *Genome Res*. 2012;22(6):1173-1183. doi:10.1101/gr.132563.111.
56. Hoque M, Ji Z, Zheng D, et al. Analysis of alterative cleavage and polyadenylation by 3′ region extraction and deep sequencing HHS Public Access. *Nat Methods*. 2013;10(2):133-139. doi:10.1038/nmeth.2288.
57. Pinto PAB, Henriques T, Freitas MO, et al. RNA polymerase II kinetics in polo polyadenylation signal selection. *EMBO J*. 2011;30(12):2431-2444. doi:10.1038/emboj.2011.156.
58. Liu X, Freitas J, Zheng D, et al. Transcription elongation rate has a tissue-specific impact on alternative cleavage and polyadenylation in Drosophila melanogaster. *RNA*. 2017;23(12):1807-1816. doi:10.1261/rna.062661.117.
59. Yeh H-S, Yong J. Alternative Polyadenylation of mRNAs: 3'-Untranslated Region Matters in Gene Expression. *Mol Cells*. 2016;39(4):281-285. doi:10.14348/molcells.2016.0035.
60. Takagaki Y, Seipelt RL, Peterson ML, Manley JL. The Polyadenylation Factor CstF-64 Regulates Alternative Processing of IgM Heavy Chain Pre-mRNA during B Cell Differentiation. *Cell*. 1996;87(5):941-952. doi:10.1016/S0092-8674(00)82000-0.
61. Duret L, Galtier N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet*. 2009;10:285-311. doi:10.1146/annurev-genom-082908-150001.
62. Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*. 2009;138(4):673-684. doi:10.1016/j.cell.2009.06.016.
63. Ji Z, Lee JY, Pan Z, Jiang B, Tian B. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci U S A*. 2009;106(17):7028-7033. doi:10.1073/pnas.0900028106.
64. Lianoglou S, Garg V, Yang JL, Leslie CS, Mayr C. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev*. 2013;27(21):2380-2396. doi:10.1101/gad.229328.113.
65. Xie X, Lu J, Kulbokas EJ, et al. Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature*. 2005;434(7031):338-345. doi:10.1038/nature03441.
66. Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*. 2008;19(1):92-105. doi:10.1101/gr.082701.108.

67. Heasman J, Wessely O, Langland R, Craig EJ, Kessler DS. Vegetal Localization of Maternal mRNAs Is Disrupted by VegT Depletion. *Dev Biol*. 2001;240(2):377-386. doi:10.1006/dbio.2001.0495.

68. Mercer TR, Wilhelm D, Dinger ME, et al. Expression of distinct RNAs from 3′ untranslated regions. *Nucleic Acids Res*. 2011;39(6):2393-2403. doi:10.1093/nar/gkq1158.

69. Kocabas A, Duarte T, Kumar S, Hynes MA. Widespread Differential Expression of Coding Region and 3′ UTR Sequences in Neurons and Other Tissues. *Neuron*. 2015;88(6):1149-1156. doi:10.1016/j.neuron.2015.10.048.

70. An JJ, Gharami K, Liao G-Y, et al. Distinct Role of Long 3′ UTR BDNF mRNA in Spine Morphology and Synaptic Plasticity in Hippocampal Neurons. *Cell*. 2008;134(1):175-187. doi:10.1016/j.cell.2008.05.045.

71. Braz SO, Cruz A, Lobo A, et al. Expression of Rac1 alternative 3′ UTRs is a cell specific mechanism with a function in dendrite outgrowth in cortical neurons. *Biochim Biophys Acta - Gene Regul Mech*. 2017;1860(6):685-694. doi:10.1016/J.BBAGRM.2017.03.002.

72. Li Y, Sun Y, Fu Y, et al. Dynamic landscape of tandem 3' UTRs during zebrafish development. *Genome Res*. 2012;22(10):1899-1906. doi:10.1101/gr.128488.111.

73. Tranter M, Helsley RN, Paulding WR, et al. Coordinated post-transcriptional regulation of Hsp70.3 gene expression by microRNA and alternative polyadenylation. *J Biol Chem*. 2011;286(34):29828-29837. doi:10.1074/jbc.M111.221796.

74. Touriol C, Morillon A, Gensac MC, Prats H, Prats AC. Expression of human fibroblast growth factor 2 mRNA is post-transcriptionally controlled by a unique destabilizing element present in the 3'-untranslated region between alternative polyadenylation sites. *J Biol Chem*. 1999;274(30):21402-21408. doi:10.1074/JBC.274.30.21402.

75. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science*. 2008;320(5883):1643-1647. doi:10.1126/science.1155390.

76. Neve J, Burger K, Li W, et al. Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. *Genome Res*. 2016;26(1):24-35. doi:10.1101/gr.193995.115.

77. Domingues RG, Lago-Baldaia I, Pereira-Castro I, et al. CD5 expression is regulated during human T-cell activation by alternative polyadenylation, PTBP1, and miR-204. *Eur J Immunol*. 2016;46(6):1490-1503. doi:10.1002/eji.201545663.

78. Spies N, Burge CB, Bartel DP. 3' UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Res*. 2013;23(12):2078-2090. doi:10.1101/gr.156919.113.

79. Gruber AR, Martin G, Müller P, et al. Global 3′ UTR shortening has a limited effect on protein abundance in proliferating T cells. *Nat Commun*. 2014;5:5465. doi:10.1038/ncomms6465.

80. Curinha A, Oliveira Braz S, Pereira-Castro I, Cruz A, Moreira A. Implications of polyadenylation in health and disease. *Nucleus*. 2014;5(6):508-519. doi:10.4161/nucl.36360.

81. Higgs DR, Goodbourn SEY, Lamb J, Clegg JB, Weatherall DJ, Proudfoot NJ. α-Thalassaemia caused by a polyadenylation signal mutation. *Nature*. 1983;306(5941):398-400. doi:10.1038/306398a0.

82. Orkin SH, Cheng TC, Antonarakis SE, Kazazian HH. Thalassemia due to a mutation in the cleavage- polyadenylation signal of the human beta- globin gene. *EMBO J*. 1985;4(2):453-456. doi:10.1002/J.1460-2075.1985.TB03650.X.

83. Gehring NH, Frede U, Neu-Yilik G, et al. Increased efficiency of mRNA 3′ end formation: a new genetic mechanism contributing to hereditary thrombophilia. *Nat Genet*. 2001;28(4):389-392. doi:10.1038/ng578.

84. Mayr C, Bartel DP. Widespread Shortening of 3′UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell*. 2009;138(4):673-684. doi:10.1016/J.CELL.2009.06.016.

85. Liaw H-H, Lin C-C, Juan H-F, Huang H-C. Differential MicroRNA Regulation Correlates with Alternative Polyadenylation Pattern between Breast Cancer and Normal Cells. Preiss T, ed. *PLoS One*. 2013;8(2):e56958. doi:10.1371/journal.pone.0056958.

86. Park HJ, Ji P, Kim S, et al. 3′ UTR shortening represses tumor-suppressor genes in trans by disrupting ceRNA crosstalk. *Nat Genet*. May 2018:1. doi:10.1038/s41588-018-0118-8.

87. Lukiw WJ, Bazan NG. Cyclooxygenase 2 RNA message abundance, stability, and hypervariability in sporadic alzheimer neocortex. *J Neurosci Res*. 1997;50(6):937-945. doi:10.1002/(SICI)1097-4547(19971215)50:6<937::AID-JNR4>3.0.CO;2-E.

88. Chen W, Jia Q, Song Y, Fu H, Wei G, Ni T. Alternative Polyadenylation: Methods, Findings, and Impacts. *Genomics Proteomics Bioinformatics*. 2017;15(5):287-300. doi:10.1016/j.gpb.2017.06.001.

89. Kozopas KM, Yang T, Buchan HL, Zhou P, Craig RW. MCL1, a gene expressed in programmed myeloid cell differentiation, has sequence similarity to BCL2. *Proc Natl Acad Sci U S A*. 1993;90(8):3516-3520. http://www.ncbi.nlm.nih.gov/pubmed/7682708. Accessed April 24, 2018.

90. Shigemasa K, Katoh O, Shiroyama Y, et al. Increased MCL-1 expression is associated with poor prognosis in ovarian carcinomas. *Jpn J Cancer Res*. 2002;93(5):542-550. http://www.ncbi.nlm.nih.gov/pubmed/12036450. Accessed May 3, 2018.

91. Campbell KJ, Dhayade S, Ferrari N, et al. MCL-1 is a prognostic indicator and drug target in breast cancer. *Cell Death Dis*. 2018;9(2):19. doi:10.1038/s41419-017-0035-2.

92. Leverson JD, Zhang H, Chen J, et al. Potent and selective small-molecule MCL-1 inhibitors demonstrate on-target cancer cell killing activity as single agents and in combination with ABT-263 (navitoclax). *Cell Death Dis*. 2015;6(1):e1590-e1590. doi:10.1038/cddis.2014.561.

93. Chen Y, Gibson SB. Mcl-1 is a Gate Keeper Regulating Cell Death in Cancer Cells. *J Clin Exp Oncol*. 2017;06(05). doi:10.4172/2324-9110.1000197.

94. Rinkenberger JL, Horning S, Klocke B, Roth K, Korsmeyer SJ. Mcl-1 deficiency results in peri-implantation embryonic lethality. *Genes Dev*. 2000;14(1):23-27. http://www.ncbi.nlm.nih.gov/pubmed/10640272. Accessed April 22, 2018.

95. Sano M, Umezawa A, Suzuki A, Shimoda K, Fukuma M, Hata J. Involvement of EAT/mcl-1, an Anti-apoptotic bcl-2-Related Gene, in Murine Embryogenesis and Human Development. *Exp Cell*

Res. 2000;259(1):127-139. doi:10.1006/EXCR.2000.4977.

96. Perciavalle RM, Opferman JT. Delving deeper: MCL-1's contributions to normal and cancer biology. *Trends Cell Biol*. 2013;23(1):22-29. doi:10.1016/j.tcb.2012.08.011.

97. Opferman JT, Iwasaki H, Ong CC, et al. Obligate Role of Anti-Apoptotic MCL-1 in the Survival of Hematopoietic Stem Cells. *Science (80-)*. 2005;307(5712):1101-1104. doi:10.1126/science.1106114.

98. Delbridge ARD, Opferman JT, Grabow S, Strasser A. Antagonism between MCL-1 and PUMA governs stem/progenitor cell survival during hematopoietic recovery from stress. *Blood*. 2015;125(21):3273-3280. doi:10.1182/blood-2015-01-621250.

99. Opferman JT, Letai A, Beard C, Sorcinelli MD, Ong CC, Korsmeyer SJ. Development and maintenance of B and T lymphocytes requires antiapoptotic MCL-1. *Nature*. 2003;426(6967):671-676. doi:10.1038/nature02067.

100. Dzhagalov I, Dunkle A, He Y-W. The anti-apoptotic Bcl-2 family member Mcl-1 promotes T lymphocyte survival at multiple stages. *J Immunol*. 2008;181(1):521-528. http://www.ncbi.nlm.nih.gov/pubmed/18566418. Accessed April 22, 2018.

101. Vikstrom I, Carotta S, Luthje K, et al. Mcl-1 Is Essential for Germinal Center Formation and B Cell Memory. *Science (80-)*. 2010;330(6007):1095-1099. doi:10.1126/science.1191793.

102. Dzhagalov I, St. John A, He Y-W. The antiapoptotic protein Mcl-1 is essential for the survival of neutrophils but not macrophages. *Blood*. 2007;109(4):1620-1626. doi:10.1182/blood-2006-03-013771.

103. Steimer DA, Boyd K, Takeuchi O, Fisher JK, Zambetti GP, Opferman JT. Selective roles for antiapoptotic MCL-1 during granulocyte development and macrophage effector function. *Blood*. 2009;113(12):2805-2815. doi:10.1182/blood-2008-05-159145.

104. Peperzak V, Vikström I, Walker J, et al. Mcl-1 is essential for the survival of plasma cells. *Nat Immunol*. 2013;14(3):290-297. doi:10.1038/ni.2527.

105. Arbour N, Vanderluit JL, Le Grand JN, et al. Mcl-1 is a key regulator of apoptosis during CNS development and after DNA damage. *J Neurosci*. 2008;28(24):6068-6078. doi:10.1523/JNEUROSCI.4940-07.2008.

106. Malone CD, Hasan SMM, Roome RB, et al. Mcl-1 regulates the survival of adult neural precursor cells. *Mol Cell Neurosci*. 2012;49(4):439-447. doi:10.1016/j.mcn.2012.02.003.

107. Wang X, Bathina M, Lynch J, et al. Deletion of MCL-1 causes lethal cardiac failure and mitochondrial dysfunction. *Genes Dev*. 2013;27(12):1351-1364. doi:10.1101/gad.215855.113.

108. Vick B, Weber A, Urbanik T, et al. Knockout of myeloid cell leukemia-1 induces liver damage and increases apoptosis susceptibility of murine hepatocytes. *Hepatology*. 2009;49(2):627-636. doi:10.1002/hep.22664.

109. Meyerovich K, Violato NM, Fukaya M, et al. MCL-1 Is a Key Antiapoptotic Protein in Human and Rodent Pancreatic β-Cells. *Diabetes*. 2017;66(9):2446-2458. doi:10.2337/db16-1252.

110. Jain R, Sheridan JM, Policheni A, et al. A critical epithelial survival axis regulated by MCL-1 maintains thymic function in mice. *Blood*. 2017;130(23):2504-2515. doi:10.1182/blood-2017-03-771576.

111. Charles A Janeway J, Travers P, Walport M, Shlomchik MJ. Generation of lymphocytes in bone marrow and thymus. 2001. https://www.ncbi.nlm.nih.gov/books/NBK27123/. Accessed April 29, 2018.
112. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Lymphocytes and the Cellular Basis of Adaptive Immunity. 2002. https://www.ncbi.nlm.nih.gov/books/NBK26921/. Accessed April 29, 2018.
113. Patel P, Chatterjee S. Innate and Adaptive Immunity. In: *Immunity and Inflammation in Health and Disease*. Elsevier; 2018:3-13. doi:10.1016/B978-0-12-805417-8.00001-9.
114. Born WK, Reardon CL, O'Brien RL. The function of γδ T cells in innate immunity. *Curr Opin Immunol*. 2006;18(1):31-38. doi:10.1016/J.COI.2005.11.007.
115. Kawai K. γδ T Cells. In: *Immunology of the Skin*. Tokyo: Springer Japan; 2016:95-111. doi:10.1007/978-4-431-55855-2_6.
116. Pierson W, Cauwe B, Policheni A, et al. Antiapoptotic Mcl-1 is critical for the survival and niche-filling capacity of Foxp3+ regulatory T cells. *Nat Immunol*. 2013;14(9):959-965. doi:10.1038/ni.2649.
117. Carrington EM, Zhan Y, Brady JL, et al. Anti-apoptotic proteins BCL-2, MCL-1 and A1 summate collectively to maintain survival of immune cell populations both in vitro and in vivo. *Cell Death Differ*. 2017;24(5):878-888. doi:10.1038/cdd.2017.30.
118. Michels J, Johnson PWM, Packham G. Mcl-1. *Int J Biochem Cell Biol*. 2005;37(2):267-271. doi:10.1016/j.biocel.2004.04.007.
119. Thomas LW, Lam C, Edwards SW. Mcl-1; the molecular regulation of protein function. *FEBS Lett*. 2010;584(14):2981-2989. doi:10.1016/j.febslet.2010.05.061.
120. Moldoveanu T, Follis AV, Kriwacki RW, Green DR. Many players in BCL-2 family affairs. *Trends Biochem Sci*. 2014;39(3):101-111. doi:10.1016/J.TIBS.2013.12.006.
121. Wang JM, Chao JR, Chen W, Kuo ML, Yen JJ, Yang-Yen HF. The antiapoptotic gene mcl-1 is up-regulated by the phosphatidylinositol 3-kinase/Akt signaling pathway through a transcription factor complex containing CREB. *Mol Cell Biol*. 1999;19(9):6195-6206. http://www.ncbi.nlm.nih.gov/pubmed/10454566. Accessed April 29, 2018.
122. Akgul C, Turner PC, White MRH, Edwards* SW. Functional analysis of the human MCL-1 gene. *Cell Mol Life Sci*. 2000;57(4):684-691. doi:10.1007/PL00000728.
123. Leu C-M, Chang C, Hu C. Epidermal growth factor (EGF) suppresses staurosporine-induced apoptosis by inducing mcl-1 via the mitogen-activated protein kinase pathway. *Oncogene*. 2000;19(13):1665-1675. doi:10.1038/sj.onc.1203452.
124. Le Gouill S, Podar K, Amiot M, et al. VEGF induces Mcl-1 up-regulation and protects multiple myeloma cells against apoptosis. *Blood*. 2004;104(9):2886-2892. doi:10.1182/blood-2004-05-1760.
125. Huang H-M, Huang C-J, Jong J, Yen -Young. Mcl-1 is a common target of stem cell factor and interleukin-5 for apoptosis prevention activity via MEK/MAPK and PI-3K/Akt pathways. http://www.bloodjournal.org/content/bloodjournal/96/5/1764.full.pdf. Accessed April 29, 2018.
126. Jourdan M, De Vos J, Mechti N, Klein B. Regulation of Bcl-2-family proteins in myeloma cells by

three myeloma survival factors: interleukin-6, interferon-alpha and insulin-like growth factor 1. *Cell Death Differ*. 2000;7(12):1244-1252. doi:10.1038/sj.cdd.4400758.

127. Puthier D, Bataille R, Amiot M. IL-6 up-regulates Mcl-1 in human myeloma cells through JAK / STAT rather than Ras / MAP kinase pathway. *Eur J Immunol*. 1999;29(12):3945-3950. doi:10.1002/(SICI)1521-4141(199912)29:12<3945::AID-IMMU3945>3.0.CO;2-O.
128. Huang HM, Huang CJ, Yen JJ. Mcl-1 is a common target of stem cell factor and interleukin-5 for apoptosis prevention activity via MEK/MAPK and PI-3K/Akt pathways. *Blood*. 2000;96(5):1764-1771. http://www.ncbi.nlm.nih.gov/pubmed/10961875. Accessed April 29, 2018.
129. Wang J-M, Lai M-Z, Yang-Yen H-F. Interleukin-3 stimulation of mcl-1 gene transcription involves activation of the PU.1 transcription factor through a p38 mitogen-activated protein kinase-dependent pathway. *Mol Cell Biol*. 2003;23(6):1896-1909. http://www.ncbi.nlm.nih.gov/pubmed/12612065. Accessed May 2, 2018.
130. Piret J-P, Minet E, Cosse J-P, et al. Hypoxia-inducible factor-1-dependent overexpression of myeloid cell factor-1 protects hypoxic cells against tert-butyl hydroperoxide-induced apoptosis. *J Biol Chem*. 2005;280(10):9336-9344. doi:10.1074/jbc.M411858200.
131. Croxton R, Ma Y, Song L, Haura EB, Cress WD. Direct repression of the Mcl-1 promoter by E2F1. *Oncogene*. 2002;21(9):1359-1369. doi:10.1038/sj.onc.1205157.
132. Cui J, Placzek WJ, Bax BAK, Bax BAK, Bax BAX. Post-Transcriptional Regulation of Anti-Apoptotic BCL2 Family Members. 2018:1-22. doi:10.3390/ijms19010308.
133. Bae J, Leo CP, Sheau Yu Hsu, Hsueh AJW. MCL-1S, a splicing variant of the antiapoptotic BCL-2 family member MCL-1, encodes a proapoptotic protein possessing only the BH3 domain. *J Biol Chem*. 2000;275(33):25255-25261. doi:10.1074/jbc.M909826199.
134. Kim JH, Sim SH, Ha HJ, Ko JJ, Lee K, Bae J. MCL-1ES, a novel variant of MCL-1, associates with MCL-1L and induces mitochondrial cell death. *FEBS Lett*. 2009;583(17):2758-2764. doi:10.1016/j.febslet.2009.08.006.
135. Filippova N, Yang X, Wang Y, et al. The RNA-binding protein HuR promotes glioma growth and treatment resistance. *Mol Cancer Res*. 2011;9(5):648-659. doi:10.1158/1541-7786.MCR-10-0325.
136. Ebner F, Sedlyarov V, Tasciyan S, et al. The RNA-binding protein tristetraprolin schedules apoptosis of pathogen-engaged neutrophils during bacterial infection. *J Clin Invest*. 2017;127(6):2051-2065. doi:10.1172/JCI80631.
137. Subramaniam D, Natarajan G, Ramalingam S, et al. Translation inhibition during cell cycle arrest and apoptosis: Mcl-1 is a novel target for RNA binding protein CUGBP2. *Am J Physiol Liver Physiol*. 2008;294(4):G1025-G1032. doi:10.1152/ajpgi.00602.2007.
138. Cui J, Placzek WJ. PTBP1 modulation of MCL1 expression regulates cellular apoptosis induced by antitubulin chemotherapeutics. *Nat Publ Gr*. 2016;23(10):1681-1690. doi:10.1038/cdd.2016.60.
139. Mott JL, Kobayashi S, Bronk SF, Gores GJ. mir-29 regulates Mcl-1 protein expression and apoptosis. *Oncogene*. 2007;26(42):6133-6140. doi:10.1038/sj.onc.1210436.
140. Mills JR, Hippo Y, Robert F, et al. mTORC1 promotes survival through translational control of Mcl-1. *Proc Natl Acad Sci U S A*. 2008;105(31):10853-10858. doi:10.1073/pnas.0804821105.

141. Tait SWG, Green DR. Mitochondria and cell death: outer membrane permeabilization and beyond. *Nat Rev Mol Cell Biol*. 2010;11(9):621-632. doi:10.1038/nrm2952.
142. Llambi F, Moldoveanu T, Tait SWG, et al. A unified model of mammalian BCL-2 protein family interactions at the mitochondria. *Mol Cell*. 2011;44(4):517-531. doi:10.1016/j.molcel.2011.10.001.
143. Chipuk JE, Green DR. How do BCL-2 proteins induce mitochondrial outer membrane permeabilization? *Trends Cell Biol*. 2008;18(4):157-164. doi:10.1016/j.tcb.2008.01.007.
144. Luna-Vargas MPA, Chipuk JE. The deadly landscape of pro-apoptotic BCL-2 proteins in the outer mitochondrial membrane. *FEBS J*. 2016;283(14):2676-2689. doi:10.1111/febs.13624.
145. Chou JJ, Li H, Salvesen GS, Yuan J, Wagner G. Solution structure of BID, an intracellular amplifier of apoptotic signaling. *Cell*. 1999;96(5):615-624. doi:10.1016/S0092-8674(00)80572-3.
146. Hinds MG, Smits C, Fredericks-Short R, et al. Bim, Bad and Bmf: intrinsically unstructured BH3-only proteins that undergo a localized conformational change upon binding to prosurvival Bcl-2 targets. *Cell Death Differ*. 2007;14(1):128-136. doi:10.1038/sj.cdd.4401934.
147. Czabotar PE, Lessene G, Strasser A, Adams JM. Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat Rev Mol Cell Biol*. 2014;15(1):49-63. doi:10.1038/nrm3722.
148. Thomas LW, Lam C, Clark RE, et al. Serine 162, an Essential Residue for the Mitochondrial Localization, Stability and Anti-Apoptotic Function of Mcl-1. *PLoS One*. 2012;7(9):1-10. doi:10.1371/journal.pone.0045088.
149. Akgul C, Moulding DA, White MR, Edwards SW. In vivo localisation and stability of human Mcl-1 using green fluorescent protein (GFP) fusion proteins. *FEBS Lett*. 2000;478(1-2):72-76. doi:10.1016/S0014-5793(00)01809-3.
150. Germain M, Duronio V. The N terminus of the anti-apoptotic BCL-2 homologue MCL-1 regulates its localization and function. *J Biol Chem*. 2007;282(44):32233-32242. doi:10.1074/jbc.M706408200.
151. Nakajima W, Hicks MA, Tanaka N, Krystal GW, Harada H. Noxa determines localization and stability of MCL-1 and consequently ABT-737 sensitivity in small cell lung cancer. *Cell Death Dis*. 2014;5(2):e1052-10. doi:10.1038/cddis.2014.6.
152. Perciavalle RM, Stewart DP, Koss B, et al. Anti-apoptotic MCL-1 localizes to the mitochondrial matrix and couples mitochondrial fusion to respiration. *Nat Cell Biol*. 2012;14(6):575-583. doi:10.1038/ncb2488.
153. Bathina M, Temirov J, Cleland MM, Pelletier S, John D. Couples Mitochondrial Fusion to Respiration. 2012;14(6):575-583. doi:10.1038/ncb2488.Anti-Apoptotic.
154. Moyzis AG, Leon L, Thomas RL, Gustafsson ÅB. Abstract 375: MCL-1 Regulates Mitochondrial Dynamics and Mitophagy. *Circ Res*. 2016;119(Suppl 1).
155. Rasmussen ML, Kline LA, Park KP, et al. A Non-apoptotic Function of MCL-1 in Promoting Pluripotency and Modulating Mitochondrial Dynamics in Stem Cells. *Stem Cell Reports*. 2018;10(3):684-692. doi:10.1016/j.stemcr.2018.01.005.
156. Fujise K, Zhang D, Liu J, Yeh ETH. Regulation of Apoptosis and Cell Cycle Progression by MCL1. *J Biol Chem*. 2000;275(50):39458-39465. doi:10.1074/jbc.M006626200.

157. JAMIL S, SOBOUTI R, HOJABRPOUR P, RAJ M, KAST J, DURONIO V. A proteolytic fragment of Mcl-1 exhibits nuclear localization and regulates cell growth by interaction with Cdk1. *Biochem J*. 2005;387(3):659-667. doi:10.1042/BJ20041596.

158. Pawlikowska P, Leray I, de Laval B, et al. ATM-dependent expression of IEX-1 controls nuclear accumulation of Mcl-1 and the DNA damage response. *Cell Death Differ*. 2010;17(11):1739-1750. doi:10.1038/cdd.2010.56.

159. Renjini AP, Titus S, Narayan P, Murali M, Kumar Jha R, Laloraya M. STAT3 and MCL-1 associate to cause a mesenchymal epithelial transition. *J Cell Sci*. 2014;127(8):1738-1750. doi:10.1242/jcs.138214.

160. Mojica FJ, Díez-Villaseñor C, Soria E, Juez G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol*. 2000;36(1):244-246. http://www.ncbi.nlm.nih.gov/pubmed/10760181. Accessed May 24, 2018.

161. Sorek R, Kunin V, Hugenholtz P. CRISPR — a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol*. 2008;6(3):181-186. doi:10.1038/nrmicro1793.

162. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8(11):2281-2308. doi:10.1038/nprot.2013.143.

163. Chylinski K, Makarova KS, Charpentier E, Koonin E V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res*. 2014;42(10):6091-6105. doi:10.1093/nar/gku241.

164. Thermes V, Grabher C, Ristoratore F, et al. I-SceI meganuclease mediates highly efficient transgenesis in fish. *Mech Dev*. 2002;118(1-2):91-98. doi:10.1016/S0925-4773(02)00218-6.

165. Epinat J-C, Arnould S, Chames P, et al. A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res*. 2003;31(11):2952-2962. doi:10.1093/nar/gkg375.

166. Silva G, Poirot L, Galetto R, et al. Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr Gene Ther*. 2011;11(1):11-27. http://www.ncbi.nlm.nih.gov/pubmed/21182466. Accessed May 26, 2018.

167. Bibikova M, Golic M, Golic KG, Carroll D. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics*. 2002;161(3):1169-1175. http://www.ncbi.nlm.nih.gov/pubmed/12136019. Accessed May 26, 2018.

168. Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol*. 2008;26(6):695-701. doi:10.1038/nbt1398.

169. Carroll D. Genome engineering with zinc-finger nucleases. *Genetics*. 2011;188(4):773-782. doi:10.1534/genetics.111.131433.

170. Li T, Huang S, Zhao X, et al. Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res*. 2011;39(14):6315-6325. doi:10.1093/nar/gkr188.

171. Liu J, Li C, Yu Z, et al. Efficient and Specific Modifications of the Drosophila Genome by Means of an Easy TALEN Strategy. *J Genet Genomics*. 2012;39(5):209-215. doi:10.1016/J.JGG.2012.04.003.

172. Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol*. 2013;14(1):49-55. doi:10.1038/nrm3486.

173. Wei C, Liu J, Yu Z, Zhang B, Gao G, Jiao R. TALEN or Cas9 – Rapid, Efficient and Specific Choices for Genome Modifications. *J Genet Genomics*. 2013;40(6):281-289. doi:10.1016/J.JGG.2013.03.013.

174. Gasiunas G, Siksnys V. RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing? *Trends Microbiol*. 2013;21(11):562-567. doi:10.1016/j.tim.2013.09.001.

175. Komor AC, Badran AH, Liu DR. Erratum: CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes (Cell (2017) 168(2) (20–36)(S0092867416314659)(10.1016/j.cell.2016.10.044)). *Cell*. 2017;169(3):559. doi:10.1016/j.cell.2017.04.005.

176. Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R. One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell*. 2013;154(6):1370-1379. doi:10.1016/J.CELL.2013.08.022.

177. Wang H, Yang H, Shivalila CS, et al. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell*. 2013;153(4):910-918. doi:10.1016/J.CELL.2013.04.025.

178. Hwang WY, Fu Y, Reyon D, et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol*. 2013;31(3):227-229. doi:10.1038/nbt.2501.

179. Friedland AE, Tzur YB, Esvelt KM, Colaiácovo MP, Church GM, Calarco JA. Heritable genome editing in C. elegans via a CRISPR-Cas9 system. *Nat Methods*. 2013;10(8):741-743. doi:10.1038/nmeth.2532.

180. Cho SW, Kim S, Kim JM, Kim J-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol*. 2013;31(3):230-232. doi:10.1038/nbt.2507.

181. Mali P, Yang L, Esvelt KM, et al. RNA-guided human genome engineering via Cas9. *Science*. 2013;339(6121):823-826. doi:10.1126/science.1232033.

182. Zagorski JW, Maser TP, Liby KT, Rockwell CE. Nrf2-Dependent and -Independent Effects of *tert* -Butylhydroquinone, CDDO-Im, and H_2O_2 in Human Jurkat T Cells as Determined by CRISPR/Cas9 Gene Editing. *J Pharmacol Exp Ther*. 2017;361(2):259-267. doi:10.1124/jpet.116.238899.

183. Zhou Y, Zhu S, Cai C, et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature*. 2014;509(7501):487-491. doi:10.1038/nature13166.

184. Deltcheva E, Chylinski K, Sharma CM, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. 2011;471(7340):602-607. doi:10.1038/nature09886.

185. Pardo B, Gómez-González B, Aguilera A. DNA Repair in Mammalian Cells. *Cell Mol Life Sci*. 2009;66(6):1039-1056. doi:10.1007/s00018-009-8740-3.

186. Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife*. 2014;3:e04766. doi:10.7554/eLife.04766.

187. Bauer DE, Canver MC, Orkin SH. Generation of Genomic Deletions in Mammalian Cell Lines via

CRISPR/Cas9. *J Vis Exp*. 2014;(83):1-10. doi:10.3791/52118.

188. Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol*. 2016;34(3):339-344. doi:10.1038/nbt.3481.
189. Chu VT, Weber T, Wefers B, et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol*. 2015;33(5):543-548. doi:10.1038/nbt.3198.
190. Shalem O, Sanjana NE, Hartenian E, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (80-)*. 2014;343(6166):84-87. doi:10.1126/science.1247005.
191. Hsu PD, Scott DA, Weinstein JA, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*. 2013;31(9):827-832. doi:10.1038/nbt.2647.
192. Li W, Teng F, Li T, Zhou Q. Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. *Nat Biotechnol*. 2013;31(8):684-686. doi:10.1038/nbt.2652.
193. Ren J, Zhang X, Liu X, et al. A versatile system for rapid multiplex genome-edited CAR T cell generation. *Oncotarget*. 2017;8(10):17002-17011. doi:10.18632/oncotarget.15218.
194. Ren J, Liu X, Fang C, Jiang S, June CH, Zhao Y. Multiplex Genome Editing to Generate Universal CAR T Cells Resistant to PD1 Inhibition. *Clin Cancer Res*. 2017;23(9):2255-2266. doi:10.1158/1078-0432.CCR-16-1300.
195. (SBI) SB. PrecisionX^{TM} Cas9 SmartNuclease System Cat. # CAS8/9xxA-1. User Manual. 2013:16.
196. CRISPR-Cas9: gene editing tool may lead to breakthrough for cancer, autism. https://www.cantechletter.com/2016/03/crispr-cas9-gene-editing-tool-may-lead-to-breakthrough-for-cancer-autism/. Accessed June 1, 2018.
197. Curinha A, Pereira-Castro I, Moreira A. Regulation of MCL1 Alternative Polyadenylation-Derived mRNA Isoforms by microRNAs in Human T Cells. [master's thesis]. Oporto, Portugal: Instituto em Investigação e Inovação em Saúde; 2014.
198. Skaar JR, Ferris AL, Wu X, et al. The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell Res*. 2015;25(3):288-305. doi:10.1038/cr.2015.19.
199. Chou C-H, Lee R-S, Yang-Yen H-F. An internal EELD domain facilitates mitochondrial targeting of Mcl-1 via a Tom70-dependent pathway. *Mol Biol Cell*. 2006;17(9):3952-3963. doi:10.1091/mbc.E06-04-0319.
200. Vasudevan S, Steitz JA. AU-Rich-Element-Mediated Upregulation of Translation by FXR1 and Argonaute 2. *Cell*. 2007;128(6):1105-1118. doi:10.1016/j.cell.2007.01.038.
201. Sureban SM, Ramalingam S, Natarajan G, et al. Translation regulatory factor RBM3 is a proto-oncogene that prevents mitotic catastrophe. *Oncogene*. 2008;27(33):4544-4556. doi:10.1038/onc.2008.97.
202. Kedde M, van Kouwenhove M, Zwart W, Oude Vrielink JAF, Elkon R, Agami R. A Pumilio-induced RNA structure switch in p27-3′ UTR controls miR-221 and miR-222 accessibility. *Nat Cell Biol*. 2010;12(10):1014-1020. doi:10.1038/ncb2105.

203. Szostak E, Gebauer F. Translational control by 3'-UTR-binding proteins. *Brief Funct Genomics*. 2013;12(1):58-65. doi:10.1093/bfgp/els056.

204. Stojanovski D, Bohnert M, Pfanner N, Laan M Van Der. Mechanisms of Protein Sorting in Mitochondria. 2015:1-18.

205. Zordan RE, Beliveau BJ, Trow JA, Craig NL, Cormack BP. Avoiding the ends: internal epitope tagging of proteins using transposon Tn7. *Genetics*. 2015;200(1):47-58. doi:10.1534/genetics.114.169482.

206. Hasse S, Hyman AA, Sarov M. TransgeneOmics – A transgenic platform for protein localization based function exploration. *Methods*. 2016;96:69-74. doi:10.1016/J.YMETH.2015.10.005.

207. Jordan ET, Collins M, Terefe J, Ugozzoli L, Rubio T. Optimizing electroporation conditions in primary and other difficult-to-transfect cells. *J Biomol Tech*. 2008;19(5):328-334. http://www.ncbi.nlm.nih.gov/pubmed/19183796. Accessed May 21, 2018.

208. Zhao N, Qi J, Zeng Z, et al. Transfecting the hard-to-transfect lymphoma/leukemia cells using a simple cationic polymer nanocomplex. *J Control Release*. 2012;159(1):104-110. doi:10.1016/j.jconrel.2012.01.007.

209. Zhang Z, Qiu S, Zhang X, Chen W. Optimized DNA electroporation for primary human T cell engineering. *BMC Biotechnol*. 2018;18(1):4. doi:10.1186/s12896-018-0419-0.

210. Mandal PK, Ferreira LMR, Collins R, et al. Efficient Ablation of Genes in Human Hematopoietic Stem and Effector Cells using CRISPR/Cas9. *Cell Stem Cell*. 2014;15(5):643-652. doi:10.1016/J.STEM.2014.10.004.

211. Chang L, Chiang C, Jiao Y, Teng L. Nonviral Transfection Methods of Efficient Gene Delivery: Micro-/Nano-Technology for Electroporation. In: ; 2016:175-218. doi:10.1142/9789813202528_0005.

212. Lundstrom K. Latest development in viral vectors for gene therapy. *Trends Biotechnol*. 2003;21(3):117-122. doi:10.1016/S0167-7799(02)00042-2.

213. Woods N-B, Muessig A, Schmidt M, et al. Lentiviral vector transduction of NOD/SCID repopulating cells results in multiple vector integrations per transduced cell: risk of insertional mutagenesis. *Blood*. 2003;101(4):1284-1289. doi:10.1182/blood-2002-07-2238.

214. Liang X, Potter J, Kumar S, et al. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol*. 2015;208:44-53. doi:10.1016/j.jbiotec.2015.04.024.

215. Shin HY, Wang C, Lee HK, et al. CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat Commun*. 2017;8(May):1-10. doi:10.1038/ncomms15464.

216. Shi Y, Di Giammartino DC, Taylor D, et al. Molecular Architecture of the Human Pre-mRNA 3′ Processing Complex. *Mol Cell*. 2009;33(3):365-376. doi:10.1016/j.molcel.2008.12.028.

217. Xia Z, Donehower LA, Cooper TA, et al. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. *Nat Commun*. 2014;5:5274. doi:10.1038/ncomms6274.

218. Martin G, Gruber AR, Keller W, Zavolan M. Genome-wide Analysis of Pre-mRNA 3′ End

Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3′ UTR Length. *Cell Rep*. 2012;1(6):753-763. doi:10.1016/j.celrep.2012.05.003.

219. Yao C, Biesinger J, Wan J, et al. Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc Natl Acad Sci*. 2012;109(46):18773-18778. doi:10.1073/pnas.1211101109.
220. Misra A, Ou J, Zhu LJ, Green MR. Global Promotion of Alternative Internal Exon Usage by mRNA 3′ End Formation Factors. *Mol Cell*. 2015;58(5):819-831. doi:10.1016/J.MOLCEL.2015.03.016.
221. Baillat D, Hakimi M-A, Näär AM, Shilatifard A, Cooch N, Shiekhattar R. Integrator, a Multiprotein Mediator of Small Nuclear RNA Processing, Associates with the C-Terminal Repeat of RNA Polymerase II. *Cell*. 2005;123(2):265-276. doi:10.1016/j.cell.2005.08.019.
222. Lai F, Gardini A, Zhang A, Shiekhattar R. Integrator mediates the biogenesis of enhancer RNAs. *Nature*. 2015;525(7569):399-403. doi:10.1038/nature14906.
223. Wang KC, Yang YW, Liu B, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*. 2011;472(7341):120-124. doi:10.1038/nature09819.
224. Ørom UA, Derrien T, Beringer M, et al. Long Noncoding RNAs with Enhancer-like Function in Human Cells. *Cell*. 2010;143(1):46-58. doi:10.1016/j.cell.2010.09.001.
225. De Santa F, Barozzi I, Mietton F, et al. A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers. Mattick JS, ed. *PLoS Biol*. 2010;8(5):e1000384. doi:10.1371/journal.pbio.1000384.
226. Kim T-K, Hemberg M, Gray JM, et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature*. 2010;465(7295):182-187. doi:10.1038/nature09033.
227. Albrecht TR, Shevtsov SP, Wu Y, et al. Integrator subunit 4 is a 'Symplekin-like' scaffold that associates with INTS9/11 to form the Integrator cleavage module. *Nucleic Acids Res*. 2018;46(8):4241-4255. doi:10.1093/nar/gky100.
228. Gardini A, Baillat D, Cesaroni M, et al. Integrator Regulates Transcriptional Initiation and Pause Release following Activation. *Mol Cell*. 2014;56(1):128-139. doi:10.1016/J.MOLCEL.2014.08.004.
229. Baillat D, Wagner EJ. Integrator: surprisingly diverse functions in gene expression. *Trends Biochem Sci*. 2015;40(5):257-264. doi:10.1016/J.TIBS.2015.03.005.
230. Didiano D, Hobert O. Molecular architecture of a miRNA-regulated 3'UTR. *RNA*. 2008;14:1297-1317. doi:10.1261/rna.1082708.in.
231. Song Y, Xu Y, Deng J, et al. CRISPR/Cas9-mediated mutation of tyrosinase (Tyr) 3′ UTR induce graying in rabbit. *Sci Rep*. 2017;7(1):1-8. doi:10.1038/s41598-017-01727-y.
232. Zhao W, Siegel D, Biton A, et al. CRISPR–Cas9-mediated functional dissection of 3′-UTRs. *Nucleic Acids Res*. 2017;(September):1-11. doi:10.1093/nar/gkx675.